NBER WORKING PAPER SERIES

ESTIMATING THE TECHNOLOGY OF COGNITIVE AND NONCOGNITIVE SKILL FORMATION

Flavio Cunha
James Heckman
Susanne Schennach

Working Paper 15664
http://www.nber.org/papers/w15664

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2010

This paper was supported by grants from the National Science Foundation (SES-0241858, SES-0099195, SES-0452089, SES-0752699); the National Institute of Child Health and Human Development (R01HD43411); the J. B. and M. K. Pritzker Foundation; the Susan Buffett Foundation; the American Bar Foundation; the Children's Initiative, a project of the Pritzker Family Foundation at the Harris School of Public Policy Studies at the University of Chicago; and PAES, supported by the Pew Foundation. We thank the editor and three anonymous referees for very helpful comments. We have also benefited from comments received from Orazio Attanasio, Gary Becker, Sarah Cattan, Philipp Eisenhauer, Miriam Gensowski, Jeffrey Grogger, Lars Hansen, Chris Hansman, Kevin Murphy, Petra Todd, Ben Williams, Ken Wolpin, and Junjian Yi, as well as from participants at the Yale Labor/Macro Conference (May 2006), University of Chicago Applications Workshop (June 2006), the New York University Applied Microeconomics Workshop (March 2008), the University of Indiana Macroeconomics Workshop (September 2008), the Applied Economics and Econometrics Seminar at the University of Western Ontario (October 2008), the Empirical Microeconomics and Econometrics Seminar at Boston College (November 2008), the IFS Conference on Structural Models of the Labour Market and Policy Analysis (November 2008), the New Economics of the Family Conference at the Milton Friedman Institute for Research in Economics (February 2009), the Econometrics Workshop at Penn State University (March 2009), the Applied Economics Workshop at University of Rochester (April 2009), the Economics Workshop at Universidad de los Andes, Bogota (May 2009), the Labor Workshop at University of Wisconsin-Madison (May 2009), the Bankard Workshop in Applied Microeconomics at the University of\ Virginia (May, 2009), the Economics Workshop at the University of Colorado-Boulder (September, 2009), and the Duke Economic Research Initiative (September, 2009). A website containing supplementary material is available at http://jenni.uchicago.edu/elast-sub. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Estimating the Technology of Cognitive and Noncognitive Skill Formation
Flavio Cunha, James Heckman, and Susanne Schennach
NBER Working Paper No. 15664
January 2010
JEL No. C31,J13

ABSTRACT

This paper formulates and estimates multistage production functions for childrens' cognitive and noncognitive skills. Skills are determined by parental environments and investments at different stages of childhood. We estimate the elasticity of substitution between investments in one period and stocks of skills in that period to assess the benefits of early investment in children compared to later remediation. We establish nonparametric identification of a general class of production technologies based on nonlinear factor models with endogenous inputs. A by-product of our approach is a framework for evaluating childhood and schooling interventions that does not rely on arbitrarily scaled test scores as outputs and recognizes the differential effects of the same bundle of skills in different tasks. Using the estimated technology, we determine optimal targeting of interventions to children with different parental and personal birth endowments. Substitutability decreases in later stages of the life cycle in the production of cognitive skills. It increases slightly in later stages of the life cycle in the production of noncognitive skills. This finding has important implications for the design of policies that target the disadvantaged. For some configurations of disadvantage and for some outcomes, the return to investments in the later stages of childhood may exceed that to investments in the early stage.

Flavio Cunha
Department of Economics
University of Pennsylvania
160 McNeil Building
3718 Locust Walk
Philadelphia, PA 19104-6297
and NBER
fcunha@sas.upenn.edu

James Heckman
Department of Economics
The University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and NBER
jjh@uchicago.edu

Susanne Schennach
University of Chicago
Department of Economics
1126 E. 59th Street
Chicago, IL 60637
smschenn@uchicago.edu

1 Introduction

A large body of research documents the importance of cognitive skills in producing social and economic success.[1] An emerging body of research establishes the parallel importance of noncognitive skills, i.e., personality, social and emotional traits.[2] Understanding the factors affecting the evolution of cognitive and noncognitive skills is important for understanding how to promote successful lives.[3]

This paper estimates the technology governing the formation of cognitive and noncognitive skills in childhood. We establish identification of general nonlinear factor models that enable us to determine the technology of skill formation. Our multistage technology captures different developmental phases in the life cycle of a child. We identify and estimate substitution parameters that determine the importance of early parental investment for subsequent lifetime achievement, and the costliness of later remediation if early investment is not undertaken.

Cunha and Heckman (2007) present a theoretical framework that organizes and interprets a large body of empirical evidence on child and animal development.[4] Cunha and Heckman (2008) estimate a linear dynamic factor model that exploits cross equation restrictions (covariance restrictions) to secure identification of a multistage technology for child investment.[5] With enough measurements relative to the number of latent skills and types of investment, it is possible to identify the latent state space dynamics generating the evolution of skills.

The linear technology used by Cunha and Heckman (2008) imposes the assumption that early and late investments are perfect substitutes over the feasible set of inputs. This paper identifies a more general nonlinear technology by extending linear state space and factor analysis to a nonlinear setting. This extension allows us to identify crucial elasticity of substitution parameters governing the trade-off between early and late investments in producing adult skills.

Drawing on the analyses of Schennach (2004a) and Hu and Schennach (2008), we es-

[1]See Herrnstein and Murray (1994), Murnane, Willett, and Levy (1995), and Cawley, Heckman, and Vytlacil (2001).

[2]See Heckman, Stixrud, and Urzua (2006), Borghans, Duckworth, Heckman, and ter Weel (2008) and the references they cite. See also the special issue of the *Journal of Human Resources* 43 (4), Fall 2008 on noncognitive skills.

[3]See Cunha, Heckman, Lochner, and Masterov (2006) and Cunha and Heckman (2007, 2009).

[4]This evidence is summarized in Knudsen, Heckman, Cameron, and Shonkoff (2006) and Heckman (2008).

[5]See Shumway and Stoffer (1982) and Watson and Engle (1983) for early discussions of such models. Amemiya and Yalcin (2001) survey the literature on nonlinear factor analysis in statistics. Our identification analysis is new. For a recent treatment of dynamic factor and related state space models see Durbin, Harvey, Koopman, and Shephard (2004) and the voluminous literature they cite.

tablish identification of the technology of skill formation. We relax the strong independence assumptions for error terms in the measurement equations that are maintained in Cunha and Heckman (2008) and Carneiro, Hansen, and Heckman (2003). The assumption of linearity of the technology in inputs that is used by Cunha and Heckman (2008) and Todd and Wolpin (2003, 2005) is not required because we allow inputs to interact in producing outputs. We generalize the factor-analytic index function models used by Carneiro, Hansen, and Heckman (2003) to allow for more general functional forms for measurement equations. We solve the problem of defining a scale for the output of childhood investments by anchoring test scores using adult outcomes of the child, which have a well-defined cardinal scale. We determine the latent variables that generate test scores by estimating how these latent variables predict adult outcomes.[6] Our approach sets the scale of test scores and latent variables in an interpretable metric. Using this metric, analysts can meaningfully interpret changes in output and conduct interpretable value-added analyses.[7] We also solve the problem of missing inputs in estimating technologies in a way that is much more general than the widely used framework of Olley and Pakes (1996) that assumes perfect proxies for latent factors. We allow for imperfect proxies and establish that measurement error is substantial in the data analyzed in this paper.

The plan of this paper is as follows. Section 2 briefly summarizes the previous literature to motivate our contribution to it. Section 3 presents our identification analysis. Section 4 discusses the data used to estimate the model, our estimation strategy, and the model estimates. Section 5 concludes.

2 A Model of Cognitive and Noncognitive Skill Formation

We analyze a model with multiple periods of childhood, $t \in \{1, 2, \ldots, T\}$, $T \geq 2$, followed by A periods of adult working life, $t \in \{T+1, T+2, \ldots, T+A\}$. The T childhood periods are divided into S stages of development, $s \in \{1, \ldots, S\}$, with $S \leq T$. Adult outcomes are produced by cognitive skills, $\theta_{C,T+1}$, and noncognitive skills, $\theta_{N,T+1}$ at the beginning of the adult years.[8] Denote parental investments at age t in child skill k by $I_{k,t}$, $k \in \{C, N\}$.

[6] Cawley, Heckman, and Vytlacil (1999) anchor test scores in earnings outcomes.

[7] Cunha and Heckman (2008) develop a class of anchoring functions invariant to affine transformations. This paper develops a more general class of monotonic transformations and presents a new analysis of joint identification of the anchoring equations and the technology of skill formation.

[8] This model generalizes the model of Becker and Tomes (1986), who assume only one period of childhood ($T = 1$) and consider one output associated with "human capital" that can be interpreted as a composite of cognitive (C) and noncognitive (N) skills. We do not model post-childhood investment.

Skills evolve in the following way. Each agent is born with initial conditions $\theta_1 = (\theta_{C,1}, \theta_{N,1})$. Family environments and genetic factors may influence these initial conditions (see Olds, 2002, and Levitt, 2003). We denote by $\theta_P = (\theta_{C,P}, \theta_{N,P})$ parental cognitive and noncognitive skills, respectively. $\theta_t = (\theta_{C,t}, \theta_{N,t})$ denotes the vector of skill stocks in period t. Let $\eta_t = (\eta_{C,t}, \eta_{N,t})$ denote shocks and/or unobserved inputs that affect the accumulation of cognitive and noncognitive skills, respectively. The technology of production of skill k in period t and developmental stage s depends on the stock of skills in period t, investment at t, $I_{k,t}$, parental skills, θ_P, shocks in period t, $\eta_{k,t}$, and the production function at stage s :

$$\theta_{k,t+1} = f_{k,s}\left(\theta_t, I_{k,t}, \theta_P, \eta_{k,t}\right), \tag{2.1}$$

for $k \in \{C, N\}$, $t \in \{1, 2, \ldots, T\}$, and $s \in \{1, \ldots, S\}$. We assume that $f_{k,s}$ is monotone increasing in its arguments, twice continuously differentiable, and concave in $I_{k,t}$. In this model, stocks of current period skills produce next period skills and affect the current period productivity of investments. Stocks of cognitive skills can promote the formation of noncognitive skills and *vice versa* because θ_t is an argument of (2.1).

Direct complementarity between the stock of skill l and the productivity of investment $I_{k,t}$ in producing skill k in period t arises if

$$\frac{\partial^2 f_{k,s}(\cdot)}{\partial I_{k,t} \partial \theta_{l,t}} > 0, \quad t \in \{1, \ldots, T\}, \quad l, k \in \{C, N\}.$$

Period t stocks of abilities and skills promote the acquisition of skills by making investment more productive. Students with greater early cognitive and noncognitive abilities are more efficient in later learning of both cognitive and noncognitive skills. The evidence from the early intervention literature suggests that the enriched early environments of the Abecedarian, Perry and Chicago Child-Parent Center (CPC) programs promoted greater efficiency in learning in high schools and reduced problem behaviors.[9]

Adult outcome j, Q_j, is produced by a combination of different skills at the beginning of period $T + 1$:

$$Q_j = g_j\left(\theta_{C,T+1}, \theta_{N,T+1}\right), \quad j \in \{1, \ldots, J\}.^{10} \tag{2.2}$$

These outcome equations capture the twin concepts that both cognitive and noncognitive

[9] See, e.g., Cunha et al. (2006); Heckman, Malofeeva, Pinto, and Savelyev (2009); Heckman, Moon, Pinto, Savelyev, and Yavitz (2010a,b), and Reynolds and Temple (2009).

[10] To focus on the main contribution of this paper, we focus on investment in children. Thus we assume that θ_{T+1} is the adult stock of skills for the rest of life contrary to the evidence reported in Borghans, Duckworth, Heckman, and ter Weel (2008). The technology could be extended to accommodate adult investment as in Ben-Porath (1967) or its generalization Heckman, Lochner, and Taber (1998)

skills matter for performance in most tasks in life and have different effects in different tasks in the labor market and in other areas of social performance. Outcomes include test scores, schooling, wages, occupational attainment, hours worked, criminal activity, and teenage pregnancy.

In this paper, we identify and estimate a *CES* version of technology (2.1) where we assume that $\theta_{C,t}$, $\theta_{N,t}$, $I_{C,t}$, $I_{N,t}$, $\theta_{C,P}$, $\theta_{N,P}$ are scalars. Outputs of skills at stage s are governed by

$$\theta_{C,t+1} = \left[\gamma_{s,C,1}\theta_{C,t}^{\phi_{s,C}} + \gamma_{s,C,2}\theta_{N,t}^{\phi_{s,C}} + \gamma_{s,C,3}I_{C,t}^{\phi_{s,C}} + \gamma_{s,C,4}\theta_{C,P}^{\phi_{s,C}} + \gamma_{s,C,5}\theta_{N,P}^{\phi_{s,C}}\right]^{\frac{1}{\phi_{s,C}}} \tag{2.3}$$

and

$$\theta_{N,t+1} = \left[\gamma_{s,N,1}\theta_{C,t}^{\phi_{s,N}} + \gamma_{s,N,2}\theta_{N,t}^{\phi_{s,N}} + \gamma_{s,N,3}I_{N,t}^{\phi_{s,N}} + \gamma_{s,N,4}\theta_{C,P}^{\phi_{s,N}} + \gamma_{s,N,5}\theta_{N,P}^{\phi_{s,N}}\right]^{\frac{1}{\phi_{s,N}}}, \tag{2.4}$$

where $\gamma_{s,k,l} \in [0,1]$, $\sum_l \gamma_{s,k,l} = 1$ for $k \in \{C,N\}$, $l \in \{1,\ldots,5\}$, $t \in \{1,\ldots,T\}$ and $s \in \{1,\ldots,S\}$. $\frac{1}{1-\phi_{s,k}}$ is the elasticity of substitution in the inputs producing $\theta_{k,t+1}$, where $\phi_{s,k} \in (-\infty, 1]$ for $k \in \{C,N\}$. It is a measure of how easy it is to compensate for low levels of stocks $\theta_{C,t}$ and $\theta_{N,t}$ inherited from the previous period with current levels of investment $I_{C,t}$ and $I_{N,t}$. For the moment, we ignore the shocks $\eta_{k,t}$ in (2.1), although they play an important role in our empirical analysis.

A *CES* specification of adult outcomes is:

$$Q_j = \left\{\rho_j \left(\theta_{C,T+1}\right)^{\phi_{Q,j}} + (1-\rho_j)\left(\theta_{N,T+1}\right)^{\phi_{Q,j}}\right\}^{\frac{1}{\phi_{Q,j}}}, \tag{2.5}$$

where $\rho_j \in [0,1]$, and $\phi_{Q,j} \in (-\infty, 1]$ for $j = 1,\ldots,J$. $\frac{1}{1-\phi_{Q,j}}$ is the elasticity of substitution across different skills in the production of outcome j. The ability of noncognitive skills to compensate for cognitive deficits in producing adult outcomes is governed by $\phi_{Q,j}$. The importance of cognition in producing output in task j is governed by the share parameter ρ_j.

To gain some insight into this model, consider a special case investigated in Cunha and Heckman (2007) where childhood lasts two periods ($T = 2$), there is one adult outcome ("human capital") so $J = 1$, and the elasticities of substitution are the same across technologies (2.3) and (2.4) and in the outcome (2.5), so $\phi_{s,C} = \phi_{s,N} = \phi_Q = \phi$ for all $s \in \{1,\ldots,S\}$. Assume that there is one investment good in each period that increases both cognitive and noncognitive skills, though not necessarily by the same amount, ($I_{k,t} \equiv I_t$, $k \in \{C,N\}$). In this case, the adult outcome is a function of investments, initial endowments, and parental characteristics and can be written as

$$Q=\left[\tau_1 I_1^{\phi}+\tau_2 I_2^{\phi}+\tau_3\theta_{C,1}^{\phi}+\tau_4\theta_{N,1}^{\phi}+\tau_5\theta_{C,P}^{\phi}+\tau_6\theta_{N,P}^{\phi}\right]^{\frac{1}{\phi}}, \tag{2.6}$$

where τ_i for $i=1,\ldots,6$ depend on the parameters of equations (2.3)–(2.5).[11] Cunha and Heckman (2007) analyze the optimal timing of investment using a special version of the technology embodied in (2.6).

Let $R(Q)=\sum_{t=2}^{A+2}\left(\frac{1}{1+r}\right)^t wQ$ denote the net present value of the child's future income computed with respect to the date of birth. Parents have resources M that they use to invest in period "1", I_1, and period "2", I_2. The objective of the parent is to maximize the net present value of the child's future income given parental resource constraints. Assuming an interior solution, that the price of investment in period "1" is one, the relative price of investment in period "2" is $\frac{1}{1+r}$, the optimal ratio of period "1" investment to period "2" investment is

$$\log\left(\frac{I_1}{I_2}\right)=\left(\frac{1}{1-\phi}\right)\left[\log\left(\frac{\tau_1}{\tau_2}\right)-\log(1+r)\right]. \tag{2.7}$$

Figure 1 plots the ratio of early to late investment as a function of τ_1/τ_2 for different values of ϕ. *Ceteris paribus*, the higher τ_1 relative to τ_2, the higher first period investment should be relative to second period investment. The parameters τ_1 and τ_2 are determined in part by the productivity of investments in producing skills, which are generated by the technology parameters $\gamma_{s,k,3}$, for $s\in\{1,2\}$ and $k\in\{C,N\}$. They also depend on the relative importance of cognitive skills, ρ, versus noncognitive skills, $1-\rho$, in producing the adult outcome Q. *Ceteris paribus*, if $\frac{\tau_1}{\tau_2}>(1+r)$, the higher the *CES* complementarity, (i.e., the lower ϕ), the greater is the ratio of optimal early to late investment. The greater r, the smaller should be the optimal ratio of early to late investment. In the limit, if investments complement each other strongly, optimality implies that they should be equal in both periods.

To see how these parameters affect the optimal ratio of early to late investment, suppose that early investment only produces cognitive skill, so that $\gamma_{1,N,3}=0$, and late investment only produces noncognitive skill, so that $\gamma_{2,C,3}=0$. In this case, the ratio $\left(\frac{\tau_1}{\tau_2}\right)$ can be expressed in terms of the technology and outcome function parameters:

$$\left(\frac{\tau_1}{\tau_2}\right)=\frac{\left(\rho\gamma_{2,C,1}+(1-\rho)\gamma_{2,N,1}\right)}{(1-\rho)}\frac{\gamma_{1,C,3}}{\gamma_{2,N,3}}.$$

For a given value of ρ (the weight placed on cognition in determining final outcomes), the ratio of early to late investment is higher the greater the ratio $\frac{\gamma_{1,C,3}}{\gamma_{2,N,3}}$. To investigate the role

[11] See Web Appendix 1 for the derivation of this expression in terms of the parameters of equations (2.3)–(2.5).

ρ plays in determining the optimal ratio of investments, assume that $\gamma_{2,C,1} \geq \gamma_{2,N,1}$, so that the stock of cognitive skill, $\theta_{C,1}$, is at least as effective in producing next period cognitive skill, $\theta_{C,2}$, as it is in producing next period noncognitive skill, $\theta_{N,2}$. Under this assumption, the higher ρ, that is, the more important cognitive skills are in producing Q, the higher the equilibrium ratio I_1/I_2. If, on the other hand, Q is more intensive in noncognitive skills, then I_1/I_2 is smaller.

This example builds intuition about the importance of the elasticity of substitution in determining the optimal timing of lifecycle investments. However, it oversimplifies the analysis of skill formation. It is implausible that the elasticity of substitution between skills in producing adult outcomes ($\frac{1}{1-\phi_Q}$) is the same as the elasticity of substitution between inputs in producing skills, and that a common elasticity of substitution governs the productivity of inputs in producing both cognitive and noncognitive skills.

Our analysis allows for multiple adult outcomes and multiple skills. We allow the elasticities of substitution governing the technologies for producing cognitive and noncognitive skills to differ at different stages of the life cycle and for both to be different from the elasticities of substitution for cognitive and noncognitive skills in producing adult outcomes. We test and reject the assumption that $\phi_{s,C} = \phi_{s,N}$ for $s \in \{1, \ldots, S\}$.

3 Identifying the Technology using Dynamic Factor Models

Identifying and estimating technology (2.1) is challenging. Both inputs and outputs can only be proxied. Measurement error in general nonlinear specifications of technology (2.1) raises serious econometric challenges. Inputs may be endogenous and the unobservables in the input equations may be correlated with unobservables in the technology equations.

This paper addresses these challenges. Specifically, we: (1) Determine how stocks of cognitive and noncognitive skills at date t affect the stocks of skills at date $t+1$, identifying both self productivity (the effects of $\theta_{N,t}$ on $\theta_{N,t+1}$, and $\theta_{C,t}$ on $\theta_{C,t+1}$) and cross productivity (the effects of $\theta_{C,t}$ on $\theta_{N,t+1}$ and the effects of $\theta_{N,t}$ on $\theta_{C,t+1}$) at each stage of the life cycle. (2) Develop a non-linear dynamic factor model where $(\theta_t, I_t, \theta_P)$ is proxied by vectors of measurements which include test scores and input measures as well as outcome measures. In our analysis, test scores and personality evaluations are indicators of latent skills. Parental inputs are indicators of latent investment. We account for measurement error in these proxies. (3) Estimate the elasticities of substitution for the technologies governing the production of cognitive and noncognitive skills. (4) Anchor the scale of test scores using adult outcome measures instead of relying on test scores as measures of output. This allows us to avoid relying on arbitrary test scores as measurements of output. Any monotonic function of a test score is still a valid test score. (5) Account for the endogeneity of parental investments when parents make child investment decisions in response to the characteristics of the child that may change over time as the child develops and as new information about the child is revealed.

Our analysis of identification proceeds in the following way. We start with a model where measurements are linear and separable in the latent variables, as in Cunha and Heckman (2008). We establish identification of the joint distribution of the latent variables without imposing conventional independence assumptions about measurement errors. With the joint distribution of latent variables in hand, we nonparametrically identify technology (2.1) given alternative assumptions about $\eta_{k,t}$. We then extend this analysis to identify nonparametric measurement and production models. We anchor the latent variables in adult outcomes to make their scales interpretable. Finally, we account for endogeneity of inputs in the technology equations and model investment behavior.

3.1 Identifying the Distribution of the Latent Variables

We use a general notation for all measurements to simplify the econometric analysis. Let $Z_{a,k,t,j}$ be the j^{th} measurement at time t on measure of type a for factor k. We have measurements on test scores and parental and teacher assessments of skills ($a = 1$), on investment ($a = 2$) and on parental endowments ($a = 3$). Each measurement has a cognitive and noncognitive component so $k \in \{C, N\}$. We initially assume that measurements are additively separable functions of the latent factors $\theta_{k,t}$ and $I_{k,t}$:

$$Z_{1,k,t,j} = \mu_{1,k,t,j} + \alpha_{1,k,t,j}\theta_{k,t} + \varepsilon_{1,k,t,j} \tag{3.1}$$

$$Z_{2,k,t,j} = \mu_{2,k,t,j} + \alpha_{2,k,t,j}I_{k,t} + \varepsilon_{2,k,t,j}, \tag{3.2}$$

$$\text{where } E(\varepsilon_{a,k,t,j}) = 0, j \in \{1, \dots, M_{a,k,t}\}, t \in \{1, \dots, T\}, k \in \{C, N\}, a \in \{1, 2\}$$

and where $\varepsilon_{a,k,t,j}$ are uncorrelated across the j.[12] Assuming that parental endowments are measured only once in period $t = 1$, we write

[12] An economic model that rationalizes the investment measurement equations in terms of family inputs is presented in Web Appendix 2.

$$Z_{3,k,1,j} = \mu_{3,k,1,j} + \alpha_{3,k,1,j}\theta_{k,P} + \varepsilon_{3,k,1,j}, ^{13,14} \tag{3.3}$$
$$E\left(\varepsilon_{3,k,1,j}\right) = 0, j \in \{1,\ldots,M_{3,k,1}\}, \text{ and } k \in \{C,N\}.$$

The $\alpha_{a,k,t,j}$ are factor loadings. The parameters and variables are defined conditional on X. To reduce the notational burden we keep X implicit. Following standard conventions in factor analysis, we set the scale of the factors by assuming $\alpha_{a,k,t,1} = 1$ and normalize $E(\theta_{k,t}) = 0$ and $E\left(I_{k,t}\right) = 0$ for all $k \in \{C,N\}, t = 1,\ldots,T$. Separability makes the identification analysis transparent. We consider a more general nonseparable model below. Given measurements $Z_{a,k,t,j}$, we can identify the mean functions $\mu_{a,k,t,j}$, $a \in \{1,2,3\}$, $t \in \{1,\ldots,T\}$, $k \in \{C,N\}$, which may depend on the X.

3.2 Identification of the Factor Loadings and of the Joint Distributions of the Latent Variables

We first establish identification of the factor loadings under the assumption that the $\varepsilon_{a,k,t,j}$ are uncorrelated across t and that the analyst has at least two measures of each type of child skills and investments in each period t, where $T \geq 2$. Without loss of generality, we focus on $\alpha_{1,C,t,j}$ and note that similar expressions can be derived for the loadings of the other latent factors.

Since $Z_{1,C,t,1}$ and $Z_{1,C,t+1,1}$ are observed, we can compute $Cov\left(Z_{1,C,t,1}, Z_{1,C,t+1,1}\right)$ from the data. Because of the normalization $\alpha_{1,C,t,1} = 1$ for all t, we obtain:

$$Cov\left(Z_{1,C,t,1}, Z_{1,C,t+1,1}\right) = Cov\left(\theta_{C,t}, \theta_{C,t+1}\right). \tag{3.4}$$

In addition, we can compute the covariance of the second measurement on cognitive skills

[13] This formulation assumes that measurements $a \in \{1,2,3\}$ proxy only one factor. This is not strictly required for identification. One can identify the correlated factor model if there is *one* measurement for each factor that depends solely on the one factor and standard normalizations and rank conditions are imposed. The other measurements can be generated by multiple factors. This follows from the analysis of Anderson and Rubin (1956) who give precise conditions for identification in factor models. Carneiro, Hansen, and Heckman (2003) consider alternative specifications. The key idea in classical factor approaches is one normalization of the factor loading for each factor in one measurement equation to set the scale of the factor and at least one measurement dedicated to each factor.

[14] In our framework, parental skills are assumed to be constant over time as a practical matter because we only observe parental skills once.

at period t with the first measurement on cognitive skills at period $t+1$:

$$Cov\left(Z_{1,C,t,2}, Z_{1,C,t+1,1}\right) = \alpha_{1,C,t,2} Cov\left(\theta_{C,t}, \theta_{C,t+1}\right). \tag{3.5}$$

If $Cov\left(\theta_{C,t}, \theta_{C,t+1}\right) \neq 0$, one can identify the loading $\alpha_{1,C,t,2}$ from the following ratio of covariances:

$$\frac{Cov\left(Z_{1,C,t,2}, Z_{1,C,t+1,1}\right)}{Cov\left(Z_{1,C,t,1}, Z_{1,C,t+1,1}\right)} = \alpha_{1,C,t,2}.$$

If there are more than two measures of cognitive skill in each period t, we can identify $\alpha_{1,C,t,j}$ for $j \in \{2, 3, \ldots, M_{1,C,t}\}$, $t \in \{1, \ldots, T\}$ up to the normalization $\alpha_{1,C,t,1} = 1$. The assumption that the $\varepsilon_{a,k,t,j}$ are uncorrelated across t is then no longer necessary. Replacing $Z_{1,C,t+1,1}$ by $Z_{a',k',t',3}$ for some (a', k', t') which may or may not be equal to $(1, C, t)$, we may proceed in the same fashion.[15] Note that the same third measurement $Z_{a',k',t',3}$ can be reused for all a, t and k implying that in the presence of serial correlation, the total number of measurements needed for identification of the factor loadings is $2L + 1$ if there are L factors.

Once the parameters $\alpha_{1,C,t,j}$ are identified, we can rewrite (3.1), assuming $\alpha_{1,C,t,j} \neq 0$, as:

$$\frac{Z_{1,C,t,j}}{\alpha_{1,C,t,j}} = \frac{\mu_{1,C,t,j}}{\alpha_{1,C,t,j}} + \theta_{C,t} + \frac{\varepsilon_{1,C,t,j}}{\alpha_{1,C,t,j}},\ j \in \{1, 2, \ldots, M_{1,C,t}\}. \tag{3.6}$$

In this form, it is clear that the known quantities $\frac{Z_{1,C,t,j}}{\alpha_{1,C,t,j}}$ play the role of repeated error-contaminated measurements of $\theta_{C,t}$. Collecting results for all $t = 1, \ldots, T$, we can identify the joint distribution of $\{\theta_{C,t}\}_{t=1}^{T}$. Proceeding in a similar fashion for all types of measurements, $a \in \{1, 2, 3\}$, on abilities $k \in \{C, N\}$, using the analysis in Schennach (2004a,b), we can identify the joint distribution of all the latent variables. Define the matrix of latent variables by θ, where

$$\theta = \left(\{\theta_{C,t}\}_{t=1}^{T}, \{\theta_{N,t}\}_{t=1}^{T}, \{I_{C,t}\}_{t=1}^{T}, \{I_{N,t}\}_{t=1}^{T}, \theta_{C,P}, \theta_{N,P}\right).$$

Thus, we can identify the joint distribution of θ, $p(\theta)$.

Although the availability of numerous indicators for each latent factor is helpful in improving the efficiency of the estimation procedure, the identification of the model can be secured (after the factor loadings are determined) if only two measurements of each latent factor are available. Since in our empirical analysis we have at least two different measurements for each latent factor, we can define, without loss of generality, the following two

[15] The idea is to write

$$\frac{Cov\left(Z_{1,C,t,2}, Z_{a',k',t',3}\right)}{Cov\left(Z_{1,C,t,1}, Z_{a',k',t',3}\right)} = \frac{\alpha_{1,C,t,2}\alpha_{a',k',t',3} Cov\left(\theta_{C,t}, \theta_{k',t'}\right)}{\alpha_{1,C,t,1}\alpha_{a',k',t',3} Cov\left(\theta_{C,t}, \theta_{k',t'}\right)} = \frac{\alpha_{1,C,t,2}}{\alpha_{1,C,t,1}} = \alpha_{1,C,t,2}$$

This only requires uncorrelatedness across different j but not across t.

vectors

$$
\begin{aligned}
W_i &= \left(\left\{\frac{Z_{1,C,t,i}}{\alpha_{1,C,t,i}}\right\}_{t=1}^{T}, \left\{\frac{Z_{1,N,t,i}}{\alpha_{1,N,t,i}}\right\}_{t=1}^{T}, \left\{\frac{Z_{2,C,t,i}}{\alpha_{2,C,t,i}}\right\}_{t=1}^{T}, \left\{\frac{Z_{2,N,t,i}}{\alpha_{2,N,t,i}}\right\}_{t=1}^{T}, \frac{Z_{3,C,1,i}}{\alpha_{3,C,1,i}}, \frac{Z_{3,N,1,i}}{\alpha_{3,N,1,i}}\right)' \\
& i \in \{1,2\}.
\end{aligned}
$$

These vectors consist of the first and the second measurements for each factor, respectively. The corresponding measurement errors are

$$
\begin{aligned}
\omega_i &= \left(\left\{\frac{\varepsilon_{1,C,t,i}}{\alpha_{1,C,t,i}}\right\}_{t=1}^{T}, \left\{\frac{\varepsilon_{1,N,t,i}}{\alpha_{1,N,t,i}}\right\}_{t=1}^{T}, \left\{\frac{\varepsilon_{2,C,t,i}}{\alpha_{2,C,t,i}}\right\}_{t=1}^{T}, \left\{\frac{\varepsilon_{2,N,t,i}}{\alpha_{2,N,t,i}}\right\}_{t=1}^{T}, \frac{\varepsilon_{3,C,1,i}}{\alpha_{3,C,1,i}}, \frac{\varepsilon_{3,N,1,i}}{\alpha_{3,N,1,i}}\right)'. \\
& i \in \{1,2\}.
\end{aligned}
$$

Identification of the distribution of θ is obtained from the following theorem. Let L denote the total number of latent factors, which in our case is $4T + 2$.

Theorem 1 *Let W_1, W_2, θ, ω_1, ω_2 be random vectors taking values in $\mathbb{R}^L$ and related through*

$$
\begin{aligned}
W_1 &= \theta + \omega_1 \\
W_2 &= \theta + \omega_2.
\end{aligned}
$$

If (i) $E[\omega_1|\theta, \omega_2] = 0$ and (ii) ω_2 is independent from θ, then the density of θ can be expressed in terms of observable quantities as:

$$
p_\theta(\theta) = (2\pi)^{-L} \int e^{-i\chi\cdot\theta} \exp\left(\int_0^\chi \frac{E\left[iW_1 e^{i\zeta\cdot W_2}\right]}{E\left[e^{i\zeta\cdot W_2}\right]} \cdot d\zeta\right) d\chi,
$$

where in this expression $i = \sqrt{-1}$, provided that all the requisite expectations exist and $E\left[e^{i\zeta\cdot W_2}\right]$ is nonvanishing. Note that the innermost integral is the integral of a vector-valued field along a continuous path joining the origin and the point $\chi \in \mathbb{R}^L$, while the outermost integral is over the whole $\mathbb{R}^L$ space. If θ does not admit a density with respect to the Lebesgue measure, $p_\theta(\theta)$ can be interpreted within the context of the theory of distributions. If some elements of θ are perfectly measured, one may simply set the corresponding elements of W_1 and W_2 to be equal. In this way, the joint distribution of mismeasured and perfectly measured variables is identified.

Proof. See Web Appendix, Part 3.1.[16]

[16] The results of Theorem 1 are sketched informally in Schennach (2004a, footnote 11).

The striking improvement in this analysis over the analysis of Cunha and Heckman (2008) is that identification can be achieved under much weaker conditions regarding measurement errors— far fewer independence assumptions are needed. The asymmetry in the analysis of ω_1 and ω_2 generalizes previous analysis which treats these terms symmetrically. It gives the analyst a more flexible toolkit for the analysis of factor models. For example, our analysis allows analysts to accommodate heteroscedasticity in the distribution of ω_1 that may depend on ω_2 and θ. It also allows for potential correlation of components within the vectors ω_1 and ω_2, thus permitting serial correlation within a given set of measurements.

The intuition for identification in this paper, as in all factor analyses, is that the signal is common to multiple measurements but the noise is not. In order to extract the noise from the signal, the disturbances have to satisfy some form of orthogonality with respect to the signal and with respect to each other. These conditions are various uncorrelatedness assumptions, conditional mean assumptions, or conditional independence assumptions. They are used in various combinations in Theorem 1, in Theorem 2 below and in other results in this paper.

3.3 The Identification of a General Measurement Error Model

In this section, we extend the previous analysis for linear factor models to consider a measurement model of the general form

$$Z_j = a_j\left(\theta, \varepsilon_j\right) \text{ for } j \in \{1, \ldots, M\}, \tag{3.7}$$

where $M \geq 3$ and where the indicator Z_j is observed while the latent factor θ and the disturbance ε_j are not. The variables Z_j, θ, and ε_j are assumed to be vectors of the same dimension. In our application, the vector of observed indicators and corresponding disturbances is

$$\begin{aligned} Z_j &= \left(\left\{Z_{1,C,t,j}\right\}_{t=1}^{T}, \left\{Z_{1,N,t,j}\right\}_{t=1}^{T}, \left\{Z_{2,C,t,j}\right\}_{t=1}^{T}, \left\{Z_{2,N,t,j}\right\}_{t=1}^{T}, Z_{3,C,1,j}, Z_{3,N,1,j}\right)' \\ \varepsilon_j &= \left(\left\{\varepsilon_{1,C,t,j}\right\}_{t=1}^{T}, \left\{\varepsilon_{1,N,t,j}\right\}_{t=1}^{T}, \left\{\varepsilon_{2,C,t,j}\right\}_{t=1}^{T}, \left\{\varepsilon_{2,N,t,j}\right\}_{t=1}^{T}, \varepsilon_{3,C,1,j}, \varepsilon_{3,C,N,1,j}\right)' \end{aligned}$$

while the vector of unobserved latent factors is:

$$\theta = \left(\left\{\theta_{C,t}\right\}_{t=1}^{T}, \left\{\theta_{N,t}\right\}_{t=1}^{T}, \left\{I_{C,t}\right\}_{t=1}^{T}, \left\{I_{N,t}\right\}_{t=1}^{T}, \theta_{C,P}, \theta_{N,P}\right)'.$$

The functions $a_j\left(\cdot,\cdot\right)$ for $j \in \{1, \ldots, M\}$ in Equations (3.7) are unknown. It is necessary to normalize one of them (e.g., $a_1\left(\cdot,\cdot\right)$) in some way to achieve identification, as established in the following theorem.

Theorem 2 *The distribution of θ in Equations (3.7) is identified under the following conditions:*

1. *The joint density of θ, Z_1, Z_2, Z_3 is bounded and so are all their marginal and conditional densities.*[17]

2. *Z_1, Z_2, Z_3 are mutually independent conditional on θ.*

3. *$p_{Z_1|Z_2}(Z_1 \mid Z_2)$ and $p_{\theta|Z_1}(\theta \mid Z_1)$ form a bounded, complete family of distributions indexed by Z_2 and Z_1, respectively.*

4. *Whenever $\theta \neq \tilde{\theta}$, $p_{Z_3|\theta}(Z_3 \mid \theta)$ and $p_{Z_3|\theta}\left(Z_3 \mid \tilde{\theta}\right)$ differ over a set of strictly positive probability.*

5. *There exists a known functional Ψ, mapping a density to a vector, that has the property that $\Psi\left[p_{Z_1|\theta}(\cdot \mid \theta)\right] = \theta$.*

Proof. See Web Appendix, Part 3.2.[18]

The proof of Theorem 2 proceeds by casting the analysis of identification as a linear algebra problem analogous to matrix diagonalization. In contrast to the standard matrix diagonalization used in linear factor analyses, we do not work with random vectors. Instead, we work with their densities. This approach offers the advantage that the problem remains linear even when the random vectors are related nonlinearly.

The conditional independence requirement of Assumption 2 is weaker than the full independence assumption traditionally made in standard linear factor models as it allows for heteroscedasticity. Assumption 3 requires θ, Z_1, Z_2 to be vectors of the same dimensions, while Assumption 4 can be satisfied even if Z_3 is a scalar. The minimum number of measurements needed for identification is therefore $2L + 1$, which is exactly the same number of measurements as in the linear, classical measurement error case.

Versions of Assumption 3 appear in the nonparametric instrumental variable literature (e.g., Newey and Powell, 2003; Darolles, Florens, and Renault, 2002). Intuitively, the requirement that $p_{Z_1|Z_2}(Z_1|Z_2)$ forms a bounded complete family requires that the density of Z_1 vary sufficiently as Z_2 varies (and similarly for $p_{\theta|Z_1}(\theta|Z_1)$).[19]

[17] This is a density with respect to the product measure of the Lebesgue measure on $\mathbb{R}^L \times \mathbb{R}^L \times \mathbb{R}^L$ and some dominating measure μ. Hence θ, Z_1, Z_2 must be continuously distributed while Z_3 may be continuous or discrete.

[18] A vector of correctly measured variables C can trivially be added to the model by including C in the list of conditioning variables for all densities in the statement of the theorem. Theorem 2 then implies that $p_{\theta|C}(\theta|C)$ is identified. Since $p_C(C)$ is identified it follows that $p_{\theta,C}(\theta, C) = p_{\theta|C}(\theta|C)p_C(C)$ is also identified.

[19] In the case of classical measurement error, bounded completeness assumptions can be phrased in terms of primitive conditions requiring nonvanishing characteristic functions of the distributions of the measurement

Assumption 4 is automatically satisfied, for instance, if θ is univariate and $a_3(\theta, \varepsilon_3)$ is strictly increasing in θ. However, it holds much more generally. Since $a_3(\theta, \varepsilon_3)$ is nonseparable, the distribution of Z_3 conditional on θ can change with θ, thus making it possible for Assumption 4 to be satisfied even if $a_3(\theta, \varepsilon_3)$ is not strictly increasing in θ.

Assumption 5 specifies how the observed Z_1 is used to determine the scale of the unobserved θ. The most common choices of the functional Ψ would be the mean, the mode, the median, or any other well-defined measure of location. This specification allows for nonclassical measurement error. One way to satisfy this assumption is to normalize $a_1(\theta, \varepsilon_1)$ to be equal to $\theta + \varepsilon_1$, where ε_1 has zero mean, median or mode. The zero mode assumption is particularly plausible for surveys where respondents face many possible wrong answers but only one correct answer. Moving the mode of the answers away from zero would therefore require a majority of respondents to misreport in exactly the same way— an unlikely scenario. Many other nonseparable functions can also satisfy this assumption. With the distribution of $p_\theta(\theta)$ in hand, we can identify the technology using the analysis presented below in Section 3.4.

Note that Theorem 2 *does not* claim that the distributions of the errors ε_j or that the functions $a_j(\cdot, \cdot)$ are identified. In fact, it is always possible to alter the distribution of ε_j and the dependence of the function $a_j(\cdot, \cdot)$ on its second argument in ways that cancel each other out, as noted in the literature on nonseparable models.[20] However, lack of identifiability of these features of the model does not prevent identification of the distribution of θ.

Nevertheless, various normalizations ensuring that the functions $a_j(\theta, \varepsilon_j)$ are fully identified are available. For example, if each element of ε_j is normalized to be uniform (or any other known distribution), the $a_j(\theta, \varepsilon_j)$ are fully identified. Other normalizations discussed in Matzkin (2003, 2007) are also possible. Alternatively, one may assume that the $a_j(\theta, \varepsilon_j)$ are separable in ε_j with zero conditional mean of ε_j given θ.[21] We invoke these assumptions when we identify the policy function for investments in Section 3.6.2 below.

The conditions justifying Theorems 1 and 2 are not nested within each other. Their different assumptions represent different trade-offs best suited for different applications. While Theorem 1 would suffice for the empirical analysis of this paper, the general result established in Theorem 2 will likely be quite useful as larger sample sizes become available.

Carneiro, Hansen, and Heckman (2003) present an analysis for nonseparable measurement equations based on a separable latent index structure, but invoke strong independence and

errors as in Mattner (1993). However, apart from this special case, very little is known about primitive conditions for bounded completeness, and research is still ongoing on this topic. See d'Haultfoeuille (2006).

[20] See Matzkin (2003, 2007).

[21] Observe that Theorem 2 covers the identifiability of the outcome (Q_j) functions (2.2) even if we supplement the model with errors $\varepsilon_j, j \in \{1, \ldots, J\}$ that satisfy the conditions of the theorem.

"identification-at-infinity" assumptions. Our approach for identifying the distribution of θ from general nonseparable measurement equations does not require these strong assumptions. Note that it also allows the θ to determine all measurements and for the θ to be freely correlated.

3.4 Nonparametric Identification of the Technology Function

Suppose that the shocks $\eta_{k,t}$ are independent over time. Below, we analyze a more general case that allows for serial dependence. Once the density of θ is known, one can identify nonseparable technology function (2.1) for $t \in \{1, \ldots, T\}$; $k \in \{C, N\}$; and $s \in \{1, \ldots, S\}$. Even if $(\theta_t, I_t, \theta_P)$ were perfectly observed, one could not separately identify the distribution of $\eta_{k,t}$ and the function $f_{k,s}$ because, without further normalizations, a change in the density of $\eta_{k,t}$ can be undone by a change in the function $f_{k,s}$.[22]

One solution to this problem is to assume that (2.1) is additively separable in $\eta_{k,t}$. Another way to avoid this ambiguity is to normalize $\eta_{k,t}$ to have a uniform density on $[0, 1]$. Any of the normalizations suggested by Matzkin (2003, 2007) could be used. Assuming $\eta_{k,t}$ is uniform $[0, 1]$, we establish that $f_{k,s}$ is nonparametrically identified, by noting that, from the knowledge of p_θ we can calculate, for any $\bar{\theta} \in \mathbb{R}$,

$$\Pr\left[\theta_{k,t+1} \leq \bar{\theta} | \theta_t, I_{k,t}, \theta_P\right] \equiv G\left(\bar{\theta} | \theta_t, I_{k,t}, \theta_P\right).$$

We identify technology (2.1) using the relationship

$$f_{k,s}\left(\theta_t, I_{k,t}, \theta_P\right) = G^{-1}\left(\eta_{k,t} | \theta_t, I_{k,t}, \theta_P\right)$$

where $G^{-1}\left(\eta_{k,t} \mid \theta_t, I_{k,t}, \theta_P\right)$ denotes the inverse of $G\left(\bar{\theta} \mid \theta_t, I_{k,t}, \theta_P\right)$ with respect to its first argument, i.e., the value $\bar{\theta}$ such that $\eta_{k,t} = G\left(\bar{\theta} \mid \theta_t, I_{k,t}, \theta_P\right)$. By construction, this operation produces a function $f_{k,s}$ that generates outcomes $\theta_{k,t+1}$ with the appropriate distribution, because a random variable is mapped into a uniformly distributed variable under the mapping defined by its own cdf.

The more traditional separable technology with zero mean disturbance, $\theta_{k,t+1} = f_{k,s}\left(\theta_t, I_{k,t}, \theta_P\right) + \eta_{k,t}$, is covered by our analysis if we define

$$f_{k,s}\left(\theta_t, I_{k,t}, \theta_P\right) \equiv E\left[\theta_{k,t+1} \mid \theta_t, I_{k,t}, \theta_P\right],$$

where the expectation is taken under the density $p_{\theta_{k,t+1}|\theta_t, I_{k,t}, \theta_P}$, which can be calculated from

[22]See, e.g, Matzkin (2003, 2007).

p_θ. The density of $\eta_{k,t}$ conditional on all variables is identified from

$$p_{\theta_{k,t+1}|\theta_t,I_{k,t},\theta_P}\left(\eta_{k,t}\mid\theta_t,I_{k,t},\theta_P\right)=p_{\theta_{k,t+1}|\theta_t,I_{k,t},\theta_P}\left(\eta_{k,t}+E\left[\theta_{k,t+1}\mid\theta_t,I_{k,t},\theta_P\right]\mid\theta_t,I_{k,t},\theta_P\right),$$

since $p_{\theta_{k,t+1}|\theta_t,I_{k,t},\theta_P}$ is known once p_θ is known. We now show how to anchor the scales of $\theta_{C,t+1}$ and $\theta_{N,t+1}$ using measures of adult outcomes.

3.5 Anchoring Skills in an Interpretable Metric

It is common in the empirical literature on child schooling and investment to measure outcomes by test scores. However, test scores are arbitrarily scaled. To gain a better understanding of the relative importance of cognitive and noncognitive skills and their interactions and the relative importance of investments at different stages of the life cycle, it is desirable to anchor skills in a common scale. In what follows, we continue to keep the conditioning on the regressors implicit.

We model the effect of period $T+1$ cognitive and noncognitive skills on adult outcomes $Z_{4,j}$, for $j\in\{1,\ldots,J\}$.[23] Suppose that there are J_1 observed outcomes that are linear functions of cognitive and noncognitive skills at the end of childhood, i.e., in period T:

$$Z_{4,j}=\mu_{4,j}+\alpha_{4,C,j}\theta_{C,T+1}+\alpha_{4,N,j}\theta_{N,T+1}+\varepsilon_{4,j},\ \text{ for } j\in\{1,\ldots,J_1\}.$$

When adult outcomes are linear and separable functions of skills, we can define the anchoring functions to be:

$$\begin{aligned} g_{C,j}\left(\theta_{C,T+1}\right) &= \mu_{4,j}+\alpha_{4,C,j}\theta_{C,T+1} \\ g_{N,j}\left(\theta_{N,T+1}\right) &= \mu_{4,j}+\alpha_{4,N,j}\theta_{N,T+1}. \end{aligned} \tag{3.8}$$

We can also anchor using nonlinear functions. One example would be an outcome produced by a latent variable $Z^*_{4,j}$, for $j\in\{J_1+1,\ldots,J\}$:

$$Z^*_{4,j}=\tilde{g}_j\left(\theta_{C,T+1},\theta_{N,T+1}\right)-\varepsilon_{4,j}.$$

Note that we do not observe $Z^*_{4,j}$, but we observe the variable $Z_{4,j}$ which is defined as:

$$Z_{4,j}=\begin{cases}1, & \text{if } \tilde{g}_j\left(\theta_{C,T+1},\theta_{N,T+1}\right)-\varepsilon_{4,j}\geq 0\\ 0, & \text{otherwise.}\end{cases}$$

[23] The $Z_{4,j}$ correspond to the Q_j of Section 2.

In this notation

$$
\begin{aligned}
\Pr\left(Z_{4,j}=1\right|\theta_{C,T+1},\theta_{N,T+1}) &= \Pr\left[\varepsilon_{4,j}\le \tilde{g}_j\left(\theta_{C,T+1},\theta_{N,T+1}\right)\right|\theta_{C,T+1},\theta_{N,T+1}] \\
&= F_{\varepsilon_{4,j}}\left[\tilde{g}_j\left(\theta_{C,T+1},\theta_{N,T+1}\right)\right|\theta_{C,T+1},\theta_{N,T+1}] \\
&= g_j\left(\theta_{C,T+1},\theta_{N,T+1}\right).
\end{aligned}
$$

Adult outcomes such as high school graduation, criminal activity, drug use, and teenage pregnancy may be represented in this fashion.

To establish identification of $g_j\left(\theta_{C,T+1},\theta_{N,T+1}\right)$ for $j\in\{J_1+1,\dots,J\}$, we include the dummy $Z_{4,j}$ in the vector θ. Assuming that the dummy $Z_{4,j}$ is measured without error, the corresponding element of the two repeated measurement vectors W_1 and W_2 are identical and equal to $Z_{4,j}$. Theorem 1 implies that the joint density of $Z_{4,j}$, $\theta_{C,t}$ and $\theta_{N,t}$ is identified. Thus, it is possible to identify $\Pr\left[Z_{4,j}=1\mid\theta_{C,T+1},\theta_{N,T+1}\right]$.

We can extract two separate "anchors" $g_{C,j}\left(\theta_{C,T+1}\right)$ and $g_{N,j}\left(\theta_{N,T+1}\right)$ from the function $g_j\left(\theta_{C,T+1},\theta_{N,T+1}\right)$, by integrating out the other variable, e.g.,

$$
\begin{aligned}
g_{C,j}\left(\theta_{C,T+1}\right) &\equiv \int g_j\left(\theta_{C,T+1},\theta_{N,T+1}\right)p_{\theta_{N,T+1}}\left(\theta_{N,T+1}\right)d\theta_{N,T+1}, \qquad (3.9)\\
g_{N,j}\left(\theta_{N,T+1}\right) &\equiv \int g_j\left(\theta_{C,T+1},\theta_{N,T+1}\right)p_{\theta_{C,T+1}}\left(\theta_{C,T+1}\right)d\theta_{C,T+1},
\end{aligned}
$$

where the marginal densities, $p_{\theta_{j,T}}\left(\theta_{N,T+1}\right)$, $j\in\{C,N\}$ are identified by applying the preceding analysis. Both $g_{C,j}(\theta_{C,T+1})$ and $g_{N,j}\left(\theta_{N,T+1}\right)$ are assumed to be strictly monotonic in their arguments.

The "anchored" skills, denoted by $\tilde{\theta}_{j,k,t}$, are defined as

$$
\tilde{\theta}_{j,k,t}=g_{k,j}\left(\theta_{k,t}\right),\ k\in\{C,N\},\quad t\in\{1,\dots,T\}.
$$

The anchored skills inherit the subscript j because different anchors generally scale the same latent variables differently.

We combine the identification of the anchoring functions with the identification of the technology function $f_{k,s}\left(\theta_t,I_{k,t},\theta_P,\eta_{k,t}\right)$ established in the previous section to prove that the technology function expressed in terms of the anchored skills — denoted by $\tilde{f}_{k,s,j}\left(\tilde{\theta}_{j,t},I_{k,t},\theta_P,\eta_{k,t}\right)$ — is also identified. To do so, redefine the technology function to be

$$
\begin{aligned}
&\tilde{f}_{k,s,j}\left(\tilde{\theta}_{j,C,t},\tilde{\theta}_{j,N,t},I_{k,t},\theta_{C,P},\theta_{N,P},\eta_{k,t}\right) \\
&\qquad\equiv g_{k,j}\left(f_{k,s}\left(g_{C,j}^{-1}\left(\tilde{\theta}_{j,C,t}\right),g_{N,j}^{-1}\left(\tilde{\theta}_{j,N,t}\right),I_{k,t},\theta_{C,P},\theta_{N,P},\eta_{k,t}\right)\right),k\in\{C,N\}
\end{aligned}
$$

where $g_{k,j}^{-1}(\cdot)$ denotes the inverse of the function $g_{k,j}(\cdot)$. Invertibility follows from the assumed monotonicity. It is straightforward to show that

$$\begin{aligned}
&\tilde{f}_{k,s,j}\left(\tilde{\theta}_{j,C,t},\tilde{\theta}_{j,N,t},I_{k,t},\theta_{C,P},\theta_{N,P},\eta_{k,t}\right) \\
&\quad=\tilde{f}_{k,s,j}\left(g_{C,j}\left(\theta_{C,t}\right),g_{N,j}\left(\theta_{N,t}\right),I_{k,t},\theta_{C,P},\theta_{N,P},\eta_{k,t}\right) \\
&\quad=g_{k,j}\left(f_{k,s}\left(g_{C,j}^{-1}\left(g_{C,j}\left(\theta_{C,t}\right)\right),g_{N,j}^{-1}\left(g_{N,j}\left(\theta_{N,t}\right)\right),I_{k,t},\theta_{C,P},\theta_{N,P},\eta_{k,t}\right)\right) \\
&\quad=g_{k,j}\left(f_{k,s}\left(\theta_{C,t},\theta_{N,t},I_{k,t},\theta_{C,P},\theta_{N,P},\eta_{k,t}\right)\right) \\
&\quad=g_{k,j}\left(\theta_{k,t+1}\right)=\tilde{\theta}_{k,j,t+1},
\end{aligned}$$

as desired. Hence, $\tilde{f}_{k,s,j}$ is the equation of motion for the anchored skills $\tilde{\theta}_{k,j,t+1}$ that is consistent with the equation of motion $f_{k,s}$ for the original skills $\theta_{k,t}$.

3.6 Accounting for Endogeneity of Parental Investment

3.6.1 Allowing for Unobserved Time-Invariant Heterogeneity

Thus far, we have maintained the assumption that the error term $\eta_{k,t}$ in the technology (2.1) is independent of all the other inputs $(\theta_t, I_{k,t}, \theta_P)$ as well as $\eta_{\ell,t}, k \neq \ell$. This implies that variables not observed by the econometrician are not used by parents to make their decisions regarding investments $I_{k,t}$. This is a very strong assumption. The availability of data on adult outcomes can be exploited to relax this assumption and allow for endogeneity of the inputs. This subsection develops an approach for a nonlinear model based on time-invariant heterogeneity.

To see how this can be done, suppose that we observe at least three adult outcomes, so that $J \geq 3$. We can then write outcomes as functions of $T+1$ skills as well as unobserved (by the economist) time-invariant heterogeneity component, π, on which parents make their investment decisions:

$$Z_{4,j}=\alpha_{4,C,j}\theta_{C,T+1}+\alpha_{4,N,j}\theta_{N,T+1}+\alpha_{4,\pi,j}\pi+\varepsilon_{4,j}, \text{ for } j\in\{1,2,\ldots,J\}.$$

We can use the analysis of section 3.2, suitably extended to allow for measurements $Z_{4,j}$, to secure identification of the factor loadings $\alpha_{4,C,j}$, $\alpha_{4,N,j}$, and $\alpha_{4,\pi,j}$. We can apply the argument of section 3.4 to secure identification of the joint distribution of $(\theta_t, I_t, \theta_P, \pi)$.[24] Write $\eta_{k,t}=(\pi,\nu_{k,t})$. Extending the preceding analysis, we can identify a more general version of the technology:

$$\theta_{k,t+1}=f_{k,s}\left(\theta_t,I_{k,t},\theta_P,\pi,\nu_{k,t}\right).$$

[24] We discuss the identification of the factor loadings in this case in Web Appendix 4.

π is permitted to be correlated with the inputs $(\theta_t, I_t, \theta_P)$ and $\nu_{k,t}$ is assumed to be independent from the vector $(\theta_t, I_t, \theta_P, \pi)$ as well as $\nu_{l,t}$ for $l \neq k$. The next subsection develops a more general approach that allows π to vary over time.

3.6.2 More General Forms of Endogeneity

This subsection relaxes the invariant heterogeneity assumption by using exclusion restrictions based on economic theory to identify the technology under more general conditions. π_t evolves over time and agents make investment decisions based on it. Define y_t as family resources in period t (e.g., income, assets, constraints). As in Section 3, we assume that suitable multiple measurements of $\left(\theta_P, \{\theta_t, I_{C,t}, I_{N,t}, y_t\}_{t=1}^{T}\right)$ are available to identify their (joint) distribution. In our application, we assume that y_t is measured without error[25] We further assume that the error term $\eta_{k,t}$ can be decomposed into two components: $(\pi_t, \nu_{k,t})$ so that we may write the technology as

$$\theta_{k,t+1} = f_{k,s}\left(\theta_t, I_{k,t}, \theta_P, \pi_t, \nu_{k,t}\right). \tag{3.10}$$

π_t is assumed to be a scalar shock independent over people but not over time. A common shock affects all technologies, but its effect may differ across technologies. The component $\nu_{k,t}$ is independent of $\theta_t, I_{k,t}, \theta_P, y_t$ and independent of $\nu_{k,t'}$ for $t' \neq t$. Its realization takes place at the end of period t, after investment choices have already been made and implemented. The shock π_t is realized before parents make investment choices, so we expect $I_{k,t}$ to respond to it. π_t is an innovation that is common to both production functions for skills, although it may have different effects on each.

We analyze a model of investment of the form

$$I_{k,t} = q_{k,t}\left(\theta_t, \theta_P, y_t, \pi_t\right), \;\; k \in \{C, N\}, t \in \{1, \ldots, T\}. \tag{3.11}$$

Equation (3.11) is the investment policy function that maps state variables for the parents, $(\theta_t, \theta_P, y_t, \pi_t)$, to the control variables $I_{k,t}$ for $k \in \{C, N\}$.[26]

Our analysis relies on the assumption that the disturbances π_t and $\nu_{k,t}$ in Equation (3.10) are both scalar, although all other variables may be vector-valued. If the disturbances π_t are *i.i.d.*, identification is straightforward. To see this, impose an innocuous normalization (e.g., assume a specific marginal distribution for π_t). Then, the relationship $I_{k,t} = q_{k,t}\left(\theta_t, \theta_P, y_t, \pi_t\right)$

[25] Thus the "multiple measurements" on y_t are all equal to each other in each period t.

[26] The assumption of a common shock across technologies produces singularity across the investment equations (3.11). This is not a serious problem because, as noted below in Section 4.2.5, we cannot distinguish cognitive investment from noncognitive investment in our data. We assume a single common investment so $q_{k,t}(\cdot) = q_t(\cdot)$ for $k \in \{C, N\}$.

can be identified along the lines of the argument of Section 3.2 or 3.3, provided, for instance, that π_t is independent from $(\theta_t, \theta_P, y_t)$.

If π_t is serially correlated, it is not plausible to assume independence between π_t and θ_t, because past values of π_t will have an impact on both current π_t and on current θ_t (via the effect of past π_t on past $I_{k,t}$). To address this problem, lagged values of income y_t can be used as instruments for θ_t (θ_P and y_t could serve as their own instruments). This approach works if π_t is independent of θ_P as well as past and present values of y_t. After normalization of the distribution of the disturbance π_t, the general nonseparable function q_t can be identified using quantile instrumental variable techniques (Chernozhukov, Imbens, and Newey, 2007), under standard assumptions in that literature, including monotonicity and completeness.[27]

Once the functions $q_{k,t}$ have been identified, one can obtain $q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t})$, the inverse of $q_{k,t}(\theta_t, \theta_P, y_t, \pi_t)$ with respect to its last argument, provided $q_{k,t}(\theta_t, \theta_P, y_t, \pi_t)$ is strictly monotone in π_t at all values of the arguments. We can then rewrite the technology function (3.11) as:

$$\theta_{k,t+1} = f_{k,s}\left(\theta_t, I_{k,t}, \theta_P, q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t}), \nu_{k,t}\right) \equiv f_{k,s}^{\mathrm{rf}}(\theta_t, I_{k,t}, \theta_P, y_t, \nu_{k,t}).$$

Again using standard nonseparable identification techniques and normalizations, one can show that the reduced form f^{rf} is identified. Instruments are unnecessary here, because the disturbance $\nu_{k,t}$ is assumed independent from all other variables. However, to identify the technology $f_{k,s}$, we need to disentangle the direct effect of $\theta_t, I_{k,t}, \theta_P$ on θ_{t+1} from their indirect effect through $\pi_t = q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t})$. To accomplish this, we exploit our knowledge of $q_{k,t}^{-1}(\theta_t, \theta_P, \pi_t, y_t)$ to write:

$$f_{k,s}(\theta_t, I_{k,t}, \theta_P, \pi_t, \nu_{k,t}) = f_{k,s}^{\mathrm{rf}}(\theta_t, I_{k,t}, \theta_P, y_t, \nu_{k,t})\,|_{y_t : q_{k,t}^{-1}(\theta_t, \theta_P, I_{k,t}, y_t) = \pi_t}$$

where, on the right-hand side, we set y_t such that the corresponding implied value of π_t matches its value on the left-hand side. This does not necessarily require $q_{k,t}^{-1}(\theta_t, \theta_P, y_t, I_{k,t})$ to be invertible with respect to y_t, since we only need one suitable value of y_t for each given $(\theta_t, \theta_P, I_{k,t}, \pi_t)$ and do not necessarily require a one-to-one mapping. By construction, the support of the distribution of y_t conditional on $\theta_t, \theta_P, I_{k,t}$, is sufficiently large to guarantee the existence of at least one solution because, for a fixed $\theta_t, I_{k,t}, \theta_P$, variations in π_t are entirely due to y_t. We present a more formal discussion of our identification strategy in Section 3.3 of the Web appendix.

In our empirical application, we make further parametric assumptions regarding $f_{k,s}$ and $q_{k,t}$, which open the way to a more convenient estimation methodology to account for

[27]Complete regularity conditions along with a proof are presented in Web Appendix 3.3.

endogeneity. The idea is to assume that the function $q_{k,t}(\theta_t, \theta_P, y_t, \pi_t)$ is parametrically specified and additively separable in π_t, so that its identification follows under standard instrumental variables conditions. Next, we replace $I_{k,t}$ by its value given by the policy function in the technology

$$\theta_{k,t+1} = f_{k,s}\left(\theta_t, q_{k,t}\left(\theta_t, \theta_P, y_t, \pi_t\right), \theta_P, \pi_t, \nu_{k,t}\right).$$

Eliminating $I_{k,t}$ solves the endogeneity problem because the two disturbances π_t and $\nu_{k,t}$ are now independent of all explanatory variables, by assumption. Identification is secured by assuming that $f_{k,s}$ is parametric and additively separable in $\nu_{k,t}$ (whose conditional mean is zero) and by assuming a parametric form for $f_{\pi_t}(\pi_t)$, the density of π_t. We can then write:

$$E\left[\theta_{k,t+1} | \theta_t, \theta_P, y_t\right] = \int f_{k,s}\left(\theta_t, q_{k,t}\left(\theta_t, \theta_P, y_t, \pi_t\right), \theta_P, \pi_t, 0\right) f_{\pi_t}(\pi_t)\, d\pi_t \equiv \tilde{f}_{k,s}\left(\theta_t, \theta_P, y_t, \beta\right).$$

The right-hand is now known up to a vector of parameters β which will be (at least) locally identified if it happens that $\partial \tilde{f}_{k,s}\left(\theta_t, \theta_P, y_t, \beta\right) / \partial \beta$ evaluated at the true value of β is a vector function of θ_t, θ_P, y_t that is linearly independent. Section 4.2.5 below describes the specific functional forms used in our application.

4 Estimating the Technology of Skill Formation

Technology (2.1) and the associated measurement systems are nonparametrically identified. However, we use parametric maximum likelihood to estimate the model and do not estimate it under the most general conditions. We do this for two reasons. First, a fully nonparametric approach is too data hungry to apply to samples of the size that we have at our disposal, because the convergence rates of nonparametric estimators are quite slow. Second, solving a high-dimensional dynamic factor model is a computationally demanding task that can only be made manageable by invoking parametric assumptions. Nonetheless, the analysis of this paper shows that in principle the parametric structure used to secure the estimates reported below is not strictly required to identify the technology. The likelihood function for the model is presented in Web Appendix 5. Web Appendix 6 describes the nonlinear filtering algorithm we use to estimate the technology. Web Appendix 7 discusses how we implement anchoring. Section 8 of the Web Appendix reports a limited Monte Carlo study of a version of the general estimation strategy discussed in Section 3.6.2.

We estimate the technology on a sample of 2207 firstborn white children from the Children of the NLSY/79 (CNLSY/79) sample. Starting in 1986, the children of the NLSY/1979 female respondents, ages 0-14, have been assessed every two years. The assessments measure cognitive ability, temperament, motor and social development, behavior problems, and self-competence of the children as well as their home environments. Data are collected via direct assessment and maternal report during home visits at every biannual wave. Section 9 of the Web Appendix discusses the measurements used to proxy investment and output. Web Appendix Tables 9-1–9-3 present summary statistics of the sample we use.[28] We estimate a model for a single child and ignore interactions among children and the allocation decisions over multiple child families.

To match the biennial data collection plan, in our empirical analysis, a period is equivalent to two years. We have eight periods distributed over two stages of development.[29] We report estimates of a variety of specifications.

Dynamic factor models allow us to exploit the wealth of measures on investment and outcomes available in the CNLSY data. They solve several problems in estimating skill formation technologies. First, there are many proxies for parental investments in children's cognitive and noncognitive development. Using a dynamic factor model, we let the data pick the best combinations of family input measures that predict levels and growth in test scores. Measured inputs that are not very informative on family investment decisions will have negligible estimated factor loadings. Second, our models help us solve the problem of missing data. Assuming that the data are missing at random, we integrate out the missing items from the sample likelihood.

In practice, we cannot empirically distinguish investments in cognitive skills from investments in noncognitive skills. Accordingly, we assume investment in period t is the same for both skills although it may have different effects on those skills. Thus we assume $I_{C,t} = I_{N,t}$ and define it as I_t.

[28]While we have rich data on home inputs, the information on schooling inputs is not so rich. Consistent with results reported in Todd and Wolpin (2005), we find that the poorly measured schooling inputs in the CNLSY are estimated to have only weak and statistically insignificant effects on outputs. Even correcting for measurement error, we find no evidence for important effects of schooling inputs on child outcomes. This finding is consistent with the Coleman Report (1966) that finds weak effects of schooling inputs on child outcomes once family characteristics are entered into an analysis. We do not report estimates of the model which include schooling inputs.

[29]The first period is age 0, the second period is ages 1-2, the third period covers ages 3-4, and so on until the eighth period in which children are 13-14 years-old. The first stage of development starts at age 0 and finishes at ages 5-6, while the second stage of development starts at ages 5-6 and finishes at ages 13-14.

4.1 Empirical Specification

We use separable measurement system (3.1). We estimate versions of the technology (2.3)-(2.4) augmented to include shocks:

$$\theta_{k,t+1} = \left[\gamma_{s,k,1}\theta_{C,t}^{\phi_{s,k}} + \gamma_{s,k,2}\theta_{N,t}^{\phi_{s,k}} + \gamma_{s,k,3}I_t^{\phi_{s,k}} + \gamma_{s,k,4}\theta_{C,P}^{\phi_{s,k}} + \gamma_{s,k,5}\theta_{N,P}^{\phi_{s,k}}\right]^{\frac{1}{\phi_{s,k}}} e^{\eta_{k,t+1}}, \qquad (4.1)$$

where $\gamma_{s,k,l} \geq 0$ and $\sum_{l=1}^{5}\gamma_{s,k,l} = 1,\ k \in \{C,N\}, t \in \{1,2\}, s \in \{1,2\}$. We assume that the innovations are normally distributed: $\eta_{k,t} \sim N\left(0, \delta_{\eta,s}^2\right)$. We further assume that the $\eta_{k,t}$ are serially independent over all t and are independent of $\eta_{\ell,t}$ for $k \neq \ell$. We assume that measurements $Z_{a,k,t,j}$ proxy the *natural logarithms* of the factors. In the text, we report only anchored results.[30] For example, for $a = 1$,

$$\begin{aligned} Z_{1,k,t,j} &= \mu_{1,k,t,j} + \alpha_{1,k,t,j}\ln\theta_{k,t} + \varepsilon_{1,k,t,j} \\ &\qquad j \in \{1,\ldots,M_{a,k,t}\}, t \in \{1,\ldots,T\}, k \in \{C,N\}. \end{aligned}$$

We use the factors (and not their logarithms) as arguments of the technology.[31] This keeps the latent factors non-negative, as is required for the definition of technology (4.1). Collect the ε terms for period t into a vector ε_t. We assume that $\varepsilon_t \sim N(0,\Lambda_t)$, where Λ_t is a diagonal matrix. We impose the condition that ε_t is independent from $\varepsilon_{t'}$ for $t \neq t'$ and all $\eta_{k,t+1}$. Define the t^{th} row of θ as θ_t^r where r stands for row. Thus

$$\ln\theta_t^r = \left(\ln\theta_{C,t}, \ln\theta_{N,t}, \ln I_t, \ln\theta_{C,P}, \ln\theta_{N,P}, \ln\pi\right).$$

Identification of this model follows as a consequence of Theorems 1 and 2 and results in Matzkin (2003, 2007). We estimate the model under different assumptions about the distribution of the factors. Under the first specification, $\ln\theta_t^r$ is normally distributed with mean zero and variance-covariance matrix Σ_t. Under the second specification, $\ln\theta_t^r$ is distributed as a mixture of $\mathcal{T}$ normals. Let $\phi\left(x;\mu_{t,\tau},\Sigma_{t,\tau}\right)$ denote the density of a normal random variable with mean $\mu_{t,\tau}$ and variance-covariance matrix $\Sigma_{t,\tau}$. The mixture of normals writes the density of $\ln\theta_t^r$ as

$$p\left(\ln\theta_t^r\right) = \sum_{\tau=1}^{\mathcal{T}}\omega_\tau\phi\left(\ln\theta_t^r;\mu_{t,\tau},\Sigma_{t,\tau}\right)$$

subject to: $\sum_{\tau=1}^{\mathcal{T}}\omega_\tau = 1$ and $\sum_{\tau=1}^{\mathcal{T}}\omega_\tau\mu_{t,\tau} = 0$.

[30] Web Appendix 11.1 compares anchored and unanchored results.

[31] We use five regressors (X) for every measurement equation: a constant, the age of the child at the assessment date, the child's gender, a dummy variable if the mother was less than 20 years-old at the time of the first birth, and a cohort dummy (one if the child was born after 1987 and zero otherwise).

Our anchored results allow us to compare the productivity of investments and stocks of different skills at different stages of the life cycle on the anchored outcome. In this paper, we mainly use completed years of education by age 19, a continuous variable, as an anchor.

4.2 Empirical Estimates

This section presents results from an extensive empirical analysis estimating the multistage technology of skill formation accounting for measurement error, non-normality of the factors, endogeneity of inputs and family investment decisions. The plan of this section is as follows. We first present baseline two stage models that anchor outcomes in terms of their effects on schooling attainment, that correct for measurement errors, and that assume that the factors are normally distributed. These models do not account for endogeneity of inputs through unobserved heterogeneity components or family investment decisions. The baseline model is far more general than what is presented in previous research on the formation of child skills that uses unanchored test scores as outcome measures and does not account for measurement error.[32]

We present evidence on the first order empirical importance of measurement error. When we do not correct for it, the estimated technology suggests that there is no effect of early investment on outcomes. Controlling for endogeneity of family inputs by accounting for unobserved heterogeneity (π), and accounting explicitly for family investment decisions has substantial effects on estimated parameters.

The following empirical regularities emerge across all models that account for measurement error.[33] Self productivity of skills is greater in the second stage than in the first stage. Noncognitive skills are cross productive for cognitive skills in the first stage of production. The cross productivity effect is weaker and less precisely determined in the second stage. There is no evidence for a cross productivity effect of cognitive skills on noncognitive skills at either stage. The estimated elasticity of substitution for inputs in cognitive skill is substantially lower in the second stage of a child's life cycle than in the first stage. For noncognitive skills, the ordering is reversed for models that control for unobserved heterogeneity (π). These estimates suggest that it is easier to redress endowment deficits that determine cognition in the first stage of a child's life cycle than in the second stage. For socioemotional (noncognitive) skills, the opposite is true. For cognitive skills, the productivity parameter associated with parental investment ($\gamma_{1,C,3}$) is greater in the first stage than in the second stage ($\gamma_{2,C,3}$). For noncognitive skills, the pattern of estimates for the productivity parameter across models is less clear cut, but there are not dramatic differences across the stages.

[32] An example is the analysis of Fryer and Levitt (2004).

[33] Estimated parameters are reported in Web Appendix 10.

For both outputs, the parameter associated with the effect of parental noncognitive skills on output is smaller at the second stage than the first stage.

Web Appendix 11 discusses the sensitivity of estimates of a one-stage two-skill model to alternative anchors and to allowing for nonnormality of the factors. For these and other estimated models which are not reported, allowing for nonnormality has only minor effects on the estimates. However, anchoring affects the estimates.[34] To facilitate computation, we use years of schooling attained as the anchor in all of the models reported in this section of the paper.[35]

4.2.1 The Baseline Specification

Table 1 presents evidence on our baseline two stage model of skill formation. Outcomes are anchored in years of schooling attained. Factors are assumed to be normally distributed and we ignore heterogeneity (π). The estimates show that for both skills, self productivity increases in the second stage. Noncognitive skills foster cognitive skills in the first stage but not in the second stage. Cognitive skills have no cross-productivity effect on noncognitive skills at either stage.[36] The productivity parameter for investment is greater in the first period than the second period for either skill. The difference across stages in the estimated parameters is dramatic for cognitive skills. The variability in the shocks is greater in the second period than in the first period. The elasticity of substitution for cognitive skills is much greater in the first period than in the second period. However, the estimated elasticity of substitution for noncognitive skills increases slightly in the second stage.

For cognitive skill production, the parental cognitive skill parameter increases in the second stage. The opposite is true for parental noncognitive skills. In producing noncognitive skills, parental cognitive skills play no role at either stage. Parental noncognitive skills play a strong role in stage 1 and a weaker role in stage 2.

4.2.2 The Empirical Importance of Measurement Error

Using our factor model, we can investigate the extent of measurement error on each measure of skill and investment in our data. To simplify the notation, we keep the conditioning on the regressors implicit and, without loss of generality, consider the measurements on cognitive

[34]Cunha and Heckman (2008) show the sensitivity of the estimates to alternative anchors for a linear model specification.

[35]The normalizations for the factors are presented in Web Appendix 10.

[36]Zero values of coefficients in this and other tables arise from the optimizer attaining a boundary of zero in the parameter space.

skills in period t. For linear measurement systems, the variance can be decomposed as follows:

$$Var\left(Z_{1,C,t,j}\right) = \alpha^2_{1,C,t,j} Var\left(\ln \theta_{C,t}\right) + Var\left(\varepsilon_{1,C,t,j}\right).$$

The fractions of the variance of $Z_{1,C,t,j}$ due to measurement error, $s^{\varepsilon}_{1,C,t,j}$, and true signal, $s^{\theta}_{1,C,t,j}$ are, respectively,

$$s^{\varepsilon}_{1,C,t,j} = \frac{Var\left(\varepsilon_{1,C,t,j}\right)}{\alpha^2_{1,C,t,j} Var\left(\ln \theta_{C,t}\right) + Var\left(\varepsilon_{1,C,t,j}\right)} \text{ (noise)}$$

and

$$s^{\theta}_{1,C,t,j} = \frac{\alpha^2_{1,C,t,j} Var\left(\ln \theta_{C,t}\right)}{\alpha^2_{1,C,t,j} Var\left(\ln \theta_{C,t}\right) + Var\left(\varepsilon_{1,C,t,j}\right)} \text{ (signal).}$$

For each measure of skill and investment used in the estimation, we construct $s^{\varepsilon}_{1,C,t,j}$ and $s^{\theta}_{1,C,t,j}$ which are reported in Table 2A. Note that the early proxies tend to have a higher fraction of observed variance due to measurement error. For example, the measure that contains the lowest true signal ratio is the MSD (Motor and Social Developments Score) at year of birth, in which less than 5% of the observed variance is signal. The proxy with the highest signal ratio is the PIAT Reading Recognition Scores at ages 5-6, for which almost 96% of the observed variance is due to the variance of the true signal. Overall, about 54% of the observed variance is associated with the cognitive skill factors $\theta_{C,t}$.

Table 2A also shows the same ratios for measures of childhood noncognitive skills. The measures of noncognitive skills tend to be lower in informational content than their cognitive counterparts. Overall, less than 40% of the observed variance is due to the variance associated with the factors for noncognitive skills. The poorest measure for noncognitive skills is the "Sociability" measure at ages 3-4, in which less than 1% of the observed variance is signal. The richest is the "BPI Headstrong" score, in which almost 62% of the observed variance is due to the variance of the signal.

Table 2A also presents the signal-noise ratio of measures of parental cognitive and noncognitive skills. Overall, measures of maternal cognitive skills tend to have a higher information content than measures of noncognitive skills. While the poorest measurement on cognitive skills has a signal ratio of almost 35%, the richest measurements on noncognitive skills are slightly above 40%.

Analogous estimates of signal and noise for our investment measures are reported in Table 2B. Investment measures are much noisier than either measure of skill. The measures for investments at earlier stages tend to be noisier than the measures at later stages. It is interesting to note that the measure "Number of Books" has a high signal-noise ratio at early years, but not in later years. At earlier years, the measure "How Often Mom Reads to the Child" has about the same informational content as "Number of Books." In later years, measures such as "Trips to the Museum" and "Attendance of Musical Performances" have higher signal-noise ratios.

These estimates suggest that it is likely to be empirically important to control for measurement error in estimating technologies of skill formation. A general pattern is that at early ages measures of skill tend to be riddled with measurement error, while the reverse is true for the measurement errors for the proxies for investment.

4.2.3 The Effect of Ignoring Measurement Error on the Estimated Technology

We now demonstrate the impact of neglecting measurement error on estimates of the technology. To make the most convincing case for the importance of measurement error, we use the least error prone proxies as determined in our estimates of Table 2.[37] We continue to assume no heterogeneity.

Not accounting for measurement error has substantial effects on the estimated technology. Comparing the estimates in Table 3 with those in Table 1, the estimated first stage investment effects are much less precisely estimated in a model that ignores measurement errors than in a model that corrects for them. In the second stage, the estimated investment effects are generally stronger. Unlike all of the specifications that control for measurement error, we estimate strong cross productivity effects of cognitive skills on noncognitive skill production. As in Table 1, there are cross productivity effects of noncognitive skills on cognitive skills at both stages although the estimated productivity parameters are somewhat smaller. The estimated elasticities of substitution for cognitive skills at both stages are comparable across the two specifications. The elasticities of substitution for noncognitive skills are substantially lower at both stages in the specification that does not control for measurement error. The error variances of the shocks are substantially larger. Parental cognitive skills are estimated to have substantial effects on childhood cognitive skills but not their noncognitive skills. This contrasts with the estimates reported in Table 1 that show strong effects of parental noncognitive skills on childhood cognitive skills in both stages, and on noncognitive skills in

[37] At birth we use Cognitive Skill: weight at birth, Noncognitive Skill: Temperament/Difficulty Scale, Parental Investment: Number of books. At ages 1–2 we use Cognitive Skill: Body Parts, Noncognitive Skill: Temperament/Difficulty Scale, Parental Investment: Number of books. At ages 3–4 we use Cognitive Skill: PPVT, Noncognitive Skill: BPI Headstrong, Parental Investment: How often mother reads to the child. At ages 5–6 to ages 13–14 we use Cognitive Skill: Reading Recognition, Noncognitive Skill: BPI Headstrong, Parental Investment: How often child is taken to musical performances. Maternal Skills are time invariant: For Maternal Cognitive Skill: ASVAB Arithmetic Reasoning, For Maternal Noncognitive Skill: Self-Esteem Item: I am a failure.

the first stage.

4.2.4 Controlling for Time-Invariant Unobserved Heterogeneity in the Estimated Technology

We next consider the effect of controlling for unobserved heterogeneity in the model, with estimates reported in Table 1. We follow the method discussed in Section 3.6.1. Doing so allows for endogeneity of the inputs. We break the error term for the technology into two parts: a time-invariant unobserved heterogeneity factor π that is correlated with the vector $(\theta_t, I_t, \theta_P)$ and an *i.i.d.* error term $\nu_{k,t}$ that is assumed to be uncorrelated with all other variables.

Table 4 shows that correcting for heterogeneity, the estimated coefficients for parental investments have a greater impact on cognitive skills at the first stage. The coefficient on parental investment in the first stage is $\gamma_{1,C,3} \cong 0.16$, while in the second stage $\gamma_{2,C,3} \cong 0.04$. The elasticity of substitution in the first stage is well above one, $\sigma_{1,C} = \frac{1}{1-0.31} \cong 1.45$, and in the second stage it is well below one, $\sigma_{2,C} \cong \frac{1}{1+1.24} \cong 0.44$. These estimates are statistically significantly different from each other and from the estimates of the elasticities of substitution $\sigma_{1,N}$ and $\sigma_{2,N}$.[38] These results suggest that early investments are important in producing cognitive skills. Consistent with the estimates reported in Table 1, noncognitive skills increase cognitive skills in the first stage, but not in the second stage. Parental cognitive and noncognitive skills affect the accumulation of childhood cognitive skills.

Panel B of Table 4 presents estimates of the technology of noncognitive skills. Note that, contrary to the estimates reported for the technology for cognitive skills, the elasticity of substitution increases slightly from the first stage to the second stage. For the early stage, $\sigma_{1,N} \cong 0.62$ while for the late stage, $\sigma_{2,N} \cong 0.65$. However, the elasticity is about 50% higher for investments in noncognitive skills for the late stage in comparison to the elasticity for investments in cognitive skills. The estimates of $\sigma_{1,N}$ and $\sigma_{2,N}$ are *not* statistically significantly different from each other, however.[39] The impact of parental investments is about the same at early and late stages ($\gamma_{1,N,3} \cong 0.06$ vs. $\gamma_{2,N,3} \cong 0.05$). Parental noncognitive skills affect the accumulation of a child's noncognitive skills both in early and late periods, but the same is not true for parental cognitive skills. The estimates in Table 4 show a strong effect of parental cognitive skills on either stage of the production of noncognitive skills.

[38] See Table 10-5 in Web Appendix 10.

[39] See Table 10-5 in Web Appendix 10.

4.2.5 A More General Approach to Solving the Problem of the Endogeneity of Inputs

This section relaxes the invariant heterogeneity assumption and reports empirical results from a more general model of time-varying heterogeneity. Our approach to estimation is motivated by the general analysis of Section 3.6.2, but, in the interest of computational tractability, we make parametric and distributional assumptions.

We augment the measurement system (3.1)–(3.3) by investment equation (3.11), which is motivated by economic theory. Our investment equation is

$$I_t = k_C \theta_{C,t} + k_N \theta_{N,t} + k_{C,P}\theta_{C,P} + k_{N,P}\theta_{N,P} + k_y y_t + \pi_t.^{40} \tag{4.2}$$

We substitute (4.2) into equations (3.2) and (3.11). We specify the income process as

$$\ln y_t = \rho_y \ln y_{t-1} + \nu_{y,t}, \tag{4.3}$$

and the equation of motion for π_t as

$$\pi_t = \rho_\pi \pi_{t-1} + \nu_{\pi,t}. \tag{4.4}$$

We assume that $\nu_{y,t} \perp\!\!\!\perp (\theta_{t'}, \nu_{y,t'})$ for all $t' \neq t$ and $\nu_{y,t} \perp\!\!\!\perp (y_{t'}, \nu_{k,t}, \theta_P)$, $t > t'$, $k \in \{C, N\}$, where "$\perp\!\!\!\perp$" means independence. We further assume that $\nu_{\pi,t} \perp\!\!\!\perp (\theta_{t'}, \theta_p, \nu_{k,t'})$ and that $(\theta_1, y_1) \perp\!\!\!\perp \pi$.[41] In addition, $\nu_{y,t} \sim N\left(0, \sigma_y^2\right)$ and $\nu_{\pi,t} \sim N\left(0, \sigma_\pi^2\right)$. In Web Appendix 8, we report favorable results from a Monte Carlo study of the estimator based on these assumptions.

Table 5 reports estimates of this model.[42] Allowing for time-varying heterogeneity does not greatly affect the estimates for fixed heterogeneity reported in Table 4. In the results that we describe below, we allow the innovation π_t to follow an AR(1) process and estimate the investment equation $q_{k,t}$ along with all of the other parameters estimated in the model reported in Table 4.[43] Estimates of the parameters of $q_{k,t}$ are presented in Web Appendix 10. We also report estimates of the anchoring equation and other outcome equations in that

[40] The intercept of the equation is absorbed into the intercept of the measurement equation.

[41] This assumption enables us to identify the parameters of equation (4.2).

[42] Table 10-6 in Web Appendix 10 reports estimates of the parameters of the investment equation (4.2).

[43] We model q as time invariant, linear and separable in its arguments, although this is not a necessary assumption in our identification, but certainly helps to save on computation time and to obtain tighter standard errors for the policy function and the production function parameters. Notice that under our assumption $I_{C,t} = I_{N,t} = I_t$, and time invariance of the investment function, it follows that $q_{k,t} = q_t = q$ for all t.

appendix.[44] When we introduce an equation for investment, the impact of early investments on the production of cognitive skill increases from $\gamma_{1,C,3} \cong 0.17$ (see Table 4, Panel A) to $\gamma_{1,C,3} \cong 0.26$ (see Table 5, Panel A). At the same time, the estimated first stage elasticity of substitution for cognitive skills increases from $\sigma_{1,C} = \frac{1}{1-\phi_{1,C}} \cong 1.5$ to $\sigma_{1,C} = \frac{1}{1-\phi_{1,C}} \cong 2.4$. Note that for this specification the impact of late investments in producing cognitive skills remains largely unchanged at $\gamma_{2,C,3} \cong 0.045$ (compare Table 4, Panel A with Table 5, Panel A). The estimate of the elasticity of substitution for cognitive skill technology falls slightly from $\sigma_{2,C} = \frac{1}{1-\phi_{2,C}} \cong 0.44$ (Table 4, Panel A) to $\sigma_{2,C} = \frac{1}{1-\phi_{2,C}} \cong 0.45$ (see Table 5, Panel A).

We obtain comparable changes in our estimates of the technology for producing noncognitive skills. The estimated impact of early investments increases from $\gamma_{1,N,3} \cong 0.05$ (see Table 4, Panel B) to $\gamma_{1,C,3} \cong 0.209$ (in Table 5, Panel B). The elasticity of substitution for noncognitive skills in the early period declines, changing from $\sigma_{2,N} = \frac{1}{1-\phi_{2,N}} \cong 0.62$ to $\sigma_{2,N} = \frac{1}{1-\phi_{2,N}} \cong 0.68$ (in Table 5, Panel B). The estimated share parameter for late investments in producing noncognitive skills increases from $\gamma_{2,C,3} \cong 0.07$ to $\gamma_{2,C,3} \cong 0.10$. Compare Table 4, Panel B with Table 5, Panel B. When we include an equation for investments, the estimated elasticity of substitution for noncognitive skills increases in late stages, from $\sigma_{2,N} = \frac{1}{1-\phi_{2,N}} \cong 0.65$ (in Table 4, Panel B) to $\sigma_{2,N} = \frac{1}{1-\phi_{2,N}} \cong 0.66$ (in Table 5, Panel B). Thus, the estimated elasticities of substitution from the more general procedure show roughly the same pattern as from the procedure that assumes time-invariant heterogeneity.[45]

The general pattern of decreasing substitution possibilities for cognitive skills and increasing substitution possibilities for noncognitive skills is consistent with the literature on the evolution of cognitive and personality traits (see Borghans et al., 2008; Shiner, 1998; Shiner and Caspi, 2003). Cognitive skills stabilize early in the life cycle. Noncognitive traits flourish, i.e., more traits are exhibited at later ages of childhood, and there are more possibilities (more margins to invest in) for compensation of disadvantage. For a more extensive discussion, see Web Appendix 1.2.

4.2.6 A Model Based Only on Cognitive Skills

Most of the empirical literature on skill production focuses on cognitive skills as the output of family investment (see, e.g., Todd and Wolpin, 2005, 2007, and the references they cite). It is of interest to estimate a more traditional model that ignores noncognitive skills and the synergism between cognitive and noncognitive skills and between investment and noncognitive skills in production. Web Appendix Table 14.1 reports estimates of a version of the

[44] We also report the covariance matrix for the initial conditions of the model in the appendix.

[45] We cannot reject the null hypothesis that $\sigma_{1,N} = \sigma_{2,N}$ but we reject the null hypothesis that $\sigma_{1,C} = \sigma_{2,C}$ and that the elasticities of different skills are equal. See Table 10-7 in Web Appendix 10.

model in Table 4 (assuming a model with time-invariant heterogeneity) where noncognitive skills are excluded from the analysis.

The estimated self-productivity effect increases from the first stage to the second stage, as occurs with the estimates found for all other specifications estimated in this paper. However, the estimated first period elasticity of substitution is much smaller than the corresponding parameter in Table 4. The estimated second period elasticity is slightly higher. The estimated productivity parameters for investment are substantially higher in both stages of the model reported in Web Appendix Table 14.1, as are the productivity parameters for parental cognitive skills. We note in the next section that the policy implications from a cognitive-skill-only model are very different from the policy implications for a model with cognitive and noncognitive skills.

4.3 Interpreting the Estimates

The major findings from our analysis of models with two skills that control for measurement error and endogeneity of inputs are: (a) Self-productivity becomes stronger as children become older, for both cognitive and noncognitive skill formation. (b) Complementarity between cognitive skills and investment becomes stronger as children become older. The elasticity of substitution for cognition is *smaller* in second stage production. It is more difficult to compensate for the effects of adverse environments on cognitive endowments at later ages than it is at earlier ages.[46] This pattern of the estimates helps to explain the evidence on ineffective cognitive remediation strategies for disadvantaged adolescents reported in Cunha, Heckman, Lochner, and Masterov (2006). (c) Complementarity between noncognitive skills and investments becomes *weaker* as children become older, but the estimated effects are not that different. The elasticity of substitution between investment and current endowments increases slightly between the first stage and the second stage in the production of noncognitive skills. It is somewhat easier at *later* stages of childhood to remediate early disadvantage using investments in noncognitive skills.

Using the estimates present in Table 4, we find that 34% of the variation in educational attainment in the sample is explained by the measures of cognitive and noncognitive capabilities that we use. 16% is due to adolescent cognitive capabilities. 12% is due to adolescent noncognitive capabilities.[47] Measured parental investments account for 15% of the variation in educational attainment. These estimates suggest that the measures of cognitive and noncognitive capabilities that we use are powerful, but not exclusive, determinants of edu-

[46]This is true even in a model that omits noncognitive skills.

[47]The skills are correlated so the marginal contributions of each skill do not add up to 34%. The decomposition used to produce these estimates is discussed in Web Appendix 12.

cational attainment and that other factors, besides the measures of family investment that we use, are at work in explaining variation in educational attainment.

To examine the implications of these estimates, we analyze a standard social planning problem that can be solved solely from knowledge of the technology of skill formation and without knowledge of parental preferences and parental access to lending markets. We determine optimal allocations of investments from a fixed budget to maximize aggregate schooling for a cohort of children. We also consider a second social planning problem that minimizes aggregate crime. Our analysis assumes that the state has full control over family investment decisions. We do not model parental investment responses to the policy. These simulations produce a measure of the investment that is needed from whatever source to achieve the specified target.

Suppose that there are H children indexed by $h \in \{1, \ldots, H\}$. Let $(\theta_{C,1,h}, \theta_{N,1,h})$ denote the initial cognitive and noncognitive skills of child h. She has parents with cognitive and noncognitive skills denoted by $\theta_{C,P,h}$ and $\theta_{N,P,h}$, respectively. Let π_h denote additional unobserved determinants of outcomes. Denote $\theta_{1,h} = (\theta_{C,1,h}, \theta_{N,1,h}, \theta_{C,P,h}, \theta_{N,P,h}, \pi_h)$ and let $F(\theta_{1,h})$ denote its distribution. We draw H people from the estimated initial distribution $F(\theta_{1,h})$. We use the estimates reported in Table 4 in this simulation. The key substitution parameters are basically the same in this model and the more general model with estimates reported in Table 5.[48] The price of investment is assumed to be the same in each period.

The social planner maximizes aggregate human capital subject to a budget constraint $B = 2H$, so that the per capita budget is 2 units of investments. We draw H children from the initial distribution $F(\theta_{1,h})$, and solve the problem of how to allocate finite resources $2H$ to maximize the average education of the cohort. Formally, the social planner maximizes aggregate schooling

$$\max \bar{S} = \frac{1}{H} \sum_{h=1}^{H} S(\theta_{C,3,h}, \theta_{N,3,h}, \pi_h),$$

subject to the aggregate budget constraint,

$$\sum_{h=1}^{H} (I_{1,h} + I_{2,h}) = 2H, \tag{4.5}$$

the technology constraint,

$$\theta_{k,t+1,h} = f_{k,t}(\theta_{C,t,h}, \theta_{N,t,h}, \theta_{C,P,h}, \theta_{N,P,h}, \pi_h) \text{ for } k \in \{C, N\} \text{ and } t \in \{1, 2\},$$

[48]Simulation from the model of Section 3.6.2 (with estimates reported in Section 4.2.5) that has time-varying child quality is considerably more complicated because of the high dimensionality of the state space. We leave this for another occasion.

and the initial endowments of the child and her family. We assume no discounting. Solving this problem, we obtain optimal early and late investments, $I_{1,h}$ and $I_{2,h}$, respectively, for each child h. An analogous social planning problem is used to minimize crime.

Figures 2 (for the child's personal endowments) and 3 (for maternal endowments) show the profiles of early (left hand side graph) and late (right hand side graph) investment as a function of child and maternal endowments. For the most disadvantaged, the optimal policy is to invest a lot in the early years. Moon (2009) shows that, in actuality, society and family together invest much more in the early years of the advantaged compared to the disadvantaged. The decline in investment by level of advantage is dramatic for early investment. Second period investment profiles are much flatter and slightly favor more advantaged children. A similar profile emerges for investments to reduce aggregate crime, which for the sake of brevity, we do not display.

Figures 4 and 5 reveal that the ratio of optimal early-to-late investment as a function of the child's personal endowments declines with advantage whether the social planner seeks to maximize educational attainment (left hand side) or to minimize aggregate crime (right hand side). A somewhat similar pattern emerges for the optimal ratio of early-to-late investment as a function of maternal endowments with one interesting twist. The optimal investment ratio is non-monotonic in the mother's cognitive skill for each level of her noncognitive skills. At very low or very high levels of maternal cognitive skills, it is better to invest relatively more in the second period than if her endowment is at the mean.

The optimal ratio of early-to-late investment depends on the desired outcome, the endowments of children and the budget. Figure 6 plots the density of the ratio of early-to-late investment for education and crime.[49] Crime is more intensive in noncognitive skill than educational attainment, which depends much more strongly on cognitive skills. Because compensation for adversity in noncognitive skills is somewhat less costly in the second period, and because of discounting of costs and concavity of the technology, it is efficient to invest relatively more in noncognitive traits in the second period.[50] The opposite is true for cognitive skills. It is optimal to weight first and second period investments in the directions indicated in the figure.

These simulations suggest that the timing and level of optimal interventions for disadvantaged children depend on the conditions of disadvantage and the nature of desired outcomes. Targeted strategies are likely to be effective especially for different targets that

[49]The optimal policy is not identical for each h and depends on $\theta_{1,h}$, which varies in the population. The crime outcome is the number of arrests. Estimates of the coefficients of the outcome equations including those for crime are reported in Web Appendix 10.

[50]This is consistent with the flourishing of noncognitive traits in later stages of the child's life cycle. See the analysis in Web Appendix 1.2.

weight cognitive and noncognitive traits differently.[51]

4.3.1 Some Economic Intuition that Explains the Simulation Results

This subsection provides an intuition for the simulation results just discussed. Given the (weak) complementarity implicit in technology (2.3) and (2.4), how is it possible to obtain our result that it is optimal to invest relatively more in the early years of the most disadvantaged? The answer hinges on the interaction between different measures of disadvantage.

Consider the following example where individuals have a single capability, θ. Suppose that there are two children, A and B, born with initial skills θ_1^A and θ_1^B, respectively. Let θ_P^A and θ_P^B denote the skills of the parents A and B, respectively. Suppose that there are two periods for investment, which we denote by periods 1 (early) and 2 (late). For each period, there is a different technology that produces skills. Assume that the technology for period one is:

$$\theta_2 = \gamma_1\theta_1 + \gamma_2 I_1 + (1-\gamma_1-\gamma_2)\,\theta_P.$$

For period two it is:

$$\theta_3 = \min\{\theta_2, I_2, \theta_P\}.$$

These patterns of complementarity are polar cases that represent, in extreme form, the empirical pattern found for cognitive skill accumulation: that substitution possibilities are greater early in life compared to later in life.

The problem of society is to choose how much to invest in child A and child B in periods 1 and 2 to maximize total aggregate skills, $\theta_3^A + \theta_3^B$, subject to the resource constraint $I_1^A + I_2^A + I_1^B + I_2^B \leq M$, where M is total resources available to the family. Formally, the problem is

$$\max\left[\begin{array}{c}\min\left\{\gamma_1\theta_1^A + \gamma_2 I_1^A + (1-\gamma_1-\gamma_2)\,\theta_P^A, I_2^A, \theta_P^A\right\} + \\ \min\left\{\gamma_1\theta_1^B + \gamma_2 I_1^B + (1-\gamma_1-\gamma_2)\,\theta_P^B, I_2^B, \theta_P^B\right\}\end{array}\right]$$
$$\text{subject to: } I_1^A + I_2^A + I_1^B + I_2^B \leq M \qquad (4.6)$$

When the resource constraint (4.6) does not bind, as it does not if M is above a certain

[51] Web Appendix 13 presents additional simulations of the model for an extreme egalitarian criterion that equalizes educational attainment across all children. We reach the same qualitative conclusions about the optimality of differentially greater investment in the early years for disadvantaged children.

threshold (determined by θ_P), optimal investments are

$$I_1^A = \frac{(\gamma_1 + \gamma_2)\,\theta_P^A - \gamma_1\theta_1^A}{\gamma_2} \qquad I_1^B = \frac{(\gamma_1 + \gamma_2)\,\theta_P^B - \gamma_1\theta_1^B}{\gamma_2}$$

$$I_2^A = \theta_P^A \qquad I_2^B = \theta_P^B$$

Notice that if child A is disadvantaged compared to B on both measures of disadvantage, ($\theta_1^A < \theta_1^B$ and $\theta_A^P < \theta_B^P$), it can happen that

$$I_1^A > I_1^B, \text{ but } I_2^A < I_2^B$$

if

$$\theta_P^A - \theta_P^B > \frac{\gamma_1}{\gamma_1 + \gamma_2}\left(\theta_1^A - \theta_1^B\right).$$

Thus, if parental endowments are less negative than the childhood endowments (scaled by $\frac{\gamma_1}{\gamma_1+\gamma_2}$), it is optimal to invest more in the early years for the disadvantaged and less in the later years. Notice that since $(1 - \gamma_1 - \gamma_2) = \gamma_P$ is the productivity parameter on θ_P in the first period technology, we can rewrite this condition as $(\theta_P^A - \theta_P^B) > \frac{\gamma_1}{1-\gamma_P}(\theta_1^A - \theta_1^B)$. The higher the self-productivity (γ_1) and the higher the parental environment productivity, γ_P, the more likely will this inequality be satisfied for any fixed level of disparity.

4.4 Implications of a One Cognitive Skill Model

Web Appendix 14.1 considers the policy implications of the social planner's problem from our estimates of a model formulated solely in terms of cognitive skills. This is the traditional focus in the analysis of educational production functions. (See, e.g., Todd and Wolpin, 2003, 2007 and Hanushek and Woessmann, 2008.) The optimal policy is to invest relatively more in the early years of the initially *advantaged*. Our estimates of two-stage and one-stage models based solely on cognitive skills would indicate that it is optimal to perpetuate initial inequality, and not to invest relatively more in disadvantaged young children.

5 Conclusion

This paper formulates and estimates a multistage model of the evolution of children's cognitive and noncognitive skills as determined by parental investments at different stages of the life cycle of children. We estimate the elasticity of substitution between contemporaneous investment and stocks of skills inherited from previous periods to determine the substitutability between early and late investments. We also determine the quantitative importance of early endowments and later investments in determining schooling attainment. We account for the proxy nature of the measures of parental inputs and of outputs and find evidence for substantial measurement error which, if not accounted for, leads to badly distorted characterizations of the technology of skill formation. We establish nonparametric identification of a wide class of nonlinear factor models which enable us to determine the technology of skill formation. We present an analysis of the identification of production technologies with endogenous missing inputs that is more general than the replacement function analysis of Olley and Pakes (1996) and allows for measurement error in the proxy variables.[52] A by-product of our approach is a framework for the evaluation of childhood interventions that avoids reliance on arbitrarily scaled test scores. We develop a nonparametric approach to this problem by anchoring test scores in adult outcomes with interpretable scales.

Using measures of parental investment and children's outcomes from the Children of the National Longitudinal Survey of Youth, we estimate the parameters governing the substitutability between early and late investments in cognitive and noncognitive skills. In our preferred empirical specification, we find much less evidence of malleability and substitutability for cognitive skills in later stages of a child's life cycle, while malleability for noncognitive skills is slightly greater at later ages. These estimates are consistent with evidence reported in Cunha, Heckman, Lochner, and Masterov (2006).

These results highlight the importance of focusing on non-cognitive skill development when designing remediation strategies for disadvantaged children. Investments in the early years are important for the formation of adult cognitive skills. Furthermore, policy simulations from the model suggest that there is no tradeoff between equity and efficiency. The optimal investment strategy to maximize aggregate schooling attainment is to target the most disadvantaged at younger ages. The optimal strategy favors later investment over early investment if the goal is to reduce crime.

Accounting for both cognitive and noncognitive skills makes a difference. An empirical model that ignores the impact of noncognitive skills on productivity and outcomes yields the opposite conclusion that an optimal policy would perpetuate initial advantages.

[52]See Heckman and Robb (1985), Heckman and Vytlacil (2007) and Matzkin (2007) for a discussion of replacement functions.

References

Amemiya, Y. and I. Yalcin (2001). Nonlinear factor analysis as a statistical method. *Statistical Science 16*(3), 275–294.

Anderson, T. and H. Rubin (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 5*, pp. 111–150. Berkeley: University of California Press.

Becker, G. S. and N. Tomes (1986, July). Human capital and the rise and fall of families. *Journal of Labor Economics 4*(3, Part 2), S1–S39.

Ben-Porath, Y. (1967, August). The production of human capital and the life cycle of earnings. *Journal of Political Economy 75*(4, Part 1), 352–365.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008, Fall). The economics and psychology of personality traits. *Journal of Human Resources 43*(4), 972–1059.

Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Cawley, J., J. J. Heckman, and E. J. Vytlacil (1999, November). On policies to reward the value added by educators. *Review of Economics and Statistics 81*(4), 720–727.

Cawley, J., J. J. Heckman, and E. J. Vytlacil (2001, September). Three observations on wages and measured cognitive ability. *Labour Economics 8*(4), 419–442.

Center for Human Resource Research (Ed.) (2004). *NLSY79 Child and Young Adult Data User's Guide*. Ohio State University, Columbus, Ohio.

Chernozhukov, V., G. W. Imbens, and W. K. Newey (2007, July). Nonparametric identification and estimation of non-separable models. *Journal of Econometrics 139*(1), 1–3.

Coleman, J. S. (1966). *Equality of Educational Opportunity.* Washington, DC: U.S. Dept. of Health, Education, and Welfare, Office of Education.

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F. and J. J. Heckman (2008, Fall). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources 43*(4), 738–782.

Cunha, F. and J. J. Heckman (2009, April). The economics and psychology of inequality and human development. *Journal of the European Economic Association 7*(2-3), 320–364. Presented as the Marshall Lecture, European Economics Association, Milan, Italy, August 29, 2008.

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam: North-Holland.

Darolles, S., J.-P. Florens, and E. Renault (2002). Nonparametric instrumental regression. Working Paper 05-2002, Centre interuniversitaire de recherche en économie quantitative, CIREQ.

d'Haultfoeuille, X. (2006). On the completeness condition in nonparametric instrumental problems. Working Paper, ENSAE, CREST-INSEE and Université de Paris I.

Durbin, J., A. C. Harvey, S. J. Koopman, and N. Shephard (2004). *State Space and Unobserved Component Models: Theory and Applications: Proceedings of a Conference in Honour of James Durbin*. New York NY: Cambridge University Press.

Fryer, R. and S. Levitt (2004, May). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics 86*(2), 447–464.

Hanushek, E. and L. Woessmann (2008, September). The role of cognitive skills in economic development. *Journal of Economic Literature 46*(3), 607–668.

Heckman, J. J. (2008, July). Schools, skills and synapses. *Economic Inquiry 46*(3), 289–324.

Heckman, J. J., L. J. Lochner, and C. Taber (1998, January). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics 1*(1), 1–58.

Heckman, J. J., L. Malofeeva, R. Pinto, and P. A. Savelyev (2009). The powerful role of noncognitive skills in explaining the effect of the Perry Preschool program. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010a, February). The rate of return to the Perry Preschool Program. *Journal of Public Economics 94*(1-2), 114–128.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010b). A reanalysis of the HighScope Perry Preschool Program. First draft, September, 2006. Forthcoming, *Quantitative Economics*.

Heckman, J. J. and R. Robb (1985). Alternative methods for evaluating the impact of interventions. In J. Heckman and B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*, Volume 10, pp. 156–245. New York: Cambridge University Press.

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effects in new environments. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 4875–5144. Amsterdam: Elsevier.

Herrnstein, R. J. and C. A. Murray (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hu, Y. and S. M. Schennach (2008, January). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*(1), 195–216.

Kniesner, T. J. and B. ter Weel (2008, Fall). Special issue on noncognitive skills and their development. *Journal of Human Resources 43*(4), 729–1059.

Knudsen, E. I., J. J. Heckman, J. Cameron, and J. P. Shonkoff (2006, July). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences 103*(27), 10155–10162.

Levitt, P. (2003, October). Structural and functional maturation of the developing primate brain. *Journal of Pediatrics 143*(4, Supplement), S35–S45.

Mattner, L. (1993, December). Some incomplete but boundedly complete location families. *Annals of Statistics 21*(4), 2158–2162.

Matzkin, R. L. (2003, September). Nonparametric estimation of nonadditive random functions. *Econometrica 71*(5), 1339–1375.

Matzkin, R. L. (2007). Nonparametric identification. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B. Amsterdam: Elsevier.

Moon, S. H. (2009). Multi-dimensional human skill formation with multi-dimensional parental investment. Unpublished manuscript, University of Chicago, Department of Economics.

Murnane, R. J., J. B. Willett, and F. Levy (1995, May). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics 77*(2), 251–266.

Newey, W. K. and J. L. Powell (2003, September). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Olds, D. L. (2002, September). Prenatal and infancy home visiting by nurses: From randomized trials to community replication. *Prevention Science 3*(2), 153–172.

Olley, G. S. and A. Pakes (1996, November). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Reynolds, A. J. and J. A. Temple (2009). Economic returns of investments in preschool education. In E. Zigler, W. Gilliam, and S. Jones (Eds.), *A Vision For Universal Preschool Education*, pp. 37–68. New York: Cambridge University Press.

Schennach, S. M. (2004a, January). Estimation of nonlinear models with measurement error. *Econometrica 72*(1), 33–75.

Schennach, S. M. (2004b). Nonparametric estimation in the presence of measurement error. *Econometric Theory 20*, 1046–1093.

Shiner, R. and A. Caspi (2003, January). Personality differences in childhood and adolescence: Measurement, development, and consequences. *Journal of Child Psychology and Psychiatry 44*(1), 2–32.

Shiner, R. L. (1998, November). How shall we speak of children's personalities in middle childhood? A preliminary taxonomy. *Psychological Bulletin 124*(3), 308–332.

Shumway, R. H. and D. S. Stoffer (1982, May). An approach to time series smoothing and forecasting using the em algorithm. *Journal of Time Series Analysis 3*(3), 253–264.

Todd, P. E. and K. I. Wolpin (2003, February). On the specification and estimation of the production function for cognitive achievement. *Economic Journal 113*(485), F3–33.

Todd, P. E. and K. I. Wolpin (2005). The production of cognitive achievement in children: Home, school and racial test score gaps. Working paper, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Todd, P. E. and K. I. Wolpin (2007, Winter). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital 1*(1), 91–136.

Watson, M. W. and R. F. Engle (1983, December). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics 23*(3), 385–400.

Table 1

Using the Factor Model to Correct for Measurement Error

Linear Anchoring on Educational Attainment (Years of Schooling)

No Unobserved Heterogeneity (π), Factors Normally Distributed

The Technology of Cognitive Skill Formation

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Self-Productivity)	$\gamma_{1,C,1}$	0.487	$\gamma_{2,C,1}$	0.902
		(0.030)		(0.014)
Current Period Noncognitive Skills (Cross-Productivity)	$\gamma_{1,C,2}$	0.083	$\gamma_{2,C,2}$	0.011
		(0.026)		(0.005)
Current Period Investments	$\gamma_{1,C,3}$	0.231	$\gamma_{2,C,3}$	0.020
		(0.024)		(0.006)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.050	$\gamma_{2,C,4}$	0.047
		(0.013)		(0.008)
Parental Noncognitive Skills	$\gamma_{1,C,5}$	0.148	$\gamma_{2,C,5}$	0.020
		(0.030)		(0.010)
Complementarity Parameter	$\phi_{1,C}$	0.611	$\phi_{2,C}$	-1.373
		(0.240)		(0.168)
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	2.569	$1/(1-\phi_{2,C})$	0.421
Variance of Shocks $\eta_{c,t}$	$\delta^2_{1,C}$	0.165	$\delta^2_{2,C}$	0.097
		(0.007)		(0.003)

The Technology of Noncognitive Skill Formation

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Cross-Productivity)	$\gamma_{1,N,1}$	0.000	$\gamma_{2,N,1}$	0.008
		(0.025)		0.010
Current Period Noncognitive Skills (Self-Productivity)	$\gamma_{1,N,2}$	0.649	$\gamma_{2,N,2}$	0.868
		(0.034)		0.011
Current Period Investments	$\gamma_{1,N,3}$	0.146	$\gamma_{2,N,3}$	0.055
		(0.027)		0.013
Parental Cognitive Skills	$\gamma_{1,N,4}$	0.022	$\gamma_{2,N,4}$	0.000
		(0.011)		0.007
Parental Noncognitive Skills	$\gamma_{1,N,5}$	0.183	$\gamma_{2,N,5}$	0.069
		(0.031)		0.017
Complementarity Parameter	$\phi_{1,N}$	-0.674	$\phi_{2,N}$	-0.695
		(0.324)		0.274
Implied Elasticity of Substitution	$1/(1-\phi_{1,N})$	0.597	$1/(1-\phi_{2,N})$	0.590
Variance of Shocks $\eta_{n,t}$	$\delta^2_{1,N}$	0.189	$\delta^2_{2,N}$	0.103
		(0.012)		0.004

Note: Standard errors in parenthesis

Table 2A

Percentage of Total Variance in Measurements due to Signal and Noise

Measurement of Child's Cognitive Skills	**%Signal**	**%Noise**
Gestation Length	0.501	0.499
Weight at Birth	0.557	0.443
Motor-Social Development at Birth	0.045	0.955
Motor-Social Development at Ages 1-2	0.275	0.725
Body Parts at Ages 1-2	0.308	0.692
Memory for Locations at Ages 1-2	0.160	0.840
Motor-Social Development at Ages 3-4	0.410	0.590
Picture Vocabulary at Ages 3-4	0.431	0.569
Picture Vocabulary at Ages 5-6	0.225	0.775
PIAT-Mathematics at Ages 5-6	0.314	0.686
PIAT-Reading Recognition at Ages 5-6	0.958	0.042
PIAT-Reading Comprehension at Ages 5-6	0.938	0.062
PIAT-Mathematics at Ages 7-8	0.465	0.535
PIAT-Reading Recognition at Ages 7-8	0.869	0.131
PIAT-Reading Comprehension at Ages 7-8	0.797	0.203
PIAT-Mathematics at Ages 9-10	0.492	0.508
PIAT-Reading Recognition at Ages 9-10	0.817	0.183
PIAT-Reading Comprehension at Ages 9-10	0.666	0.334
PIAT-Mathematics at Ages 11-12	0.516	0.484
PIAT-Reading Recognition at Ages 11-12	0.781	0.219
PIAT-Reading Comprehension at Ages 11-12	0.614	0.386
PIAT-Mathematics at Ages 13-14	0.537	0.463
PIAT-Reading Recognition at Ages 13-14	0.735	0.265
PIAT-Reading Comprehension at Ages 13-14	0.549	0.451
Measurement of Maternal Cognitive Skills		
ASVAB Arithmetic Reasoning	0.728	0.272
ASVAB Word Knowledge	0.625	0.375
ASVAB Paragraph Composition	0.576	0.424
ASVAB Numerical Operations	0.461	0.539
ASVAB Coding Speed	0.353	0.647
ASVAB Mathematical Knowledge	0.662	0.338
Measurement of Maternal Noncognitive Skills		
Self-Esteem "I am a person of worth"	0.277	0.723
Self-Esteem " I have good qualities"	0.349	0.651
Self-Esteem "I am a failure"	0.444	0.556
Self-Esteem "I have nothing to be proud of"	0.375	0.625
Self-Esteem "I have a positive attitude"	0.406	0.594
Self-Esteem "I wish I had more self-respect"	0.341	0.659
Self-Esteem "I feel useless at times"	0.293	0.707
Self-Esteem "I sometimes think I am no good"	0.375	0.625
Locus of Control "I have no control"	0.047	0.953
Locus of Control "I make no plans for the future"	0.064	0.936
Locus of Control "Luck is big factor in life"	0.041	0.959
Locus of Control "Luck plays big role in my life"	0.020	0.980

Measurement of Child's Noncognitive Skills	**%Signal**	**%Noise**
Difficulty at Birth	0.151	0.849
Friendliness at Birth	0.165	0.835
Compliance at Ages 1-2	0.232	0.768
Insecure at Ages 1-2	0.080	0.920
Sociability at Ages 1-2	0.075	0.925
Difficulty at Ages 1-2	0.382	0.618
Friendliness at Ages 1-2	0.189	0.811
Compliance at Ages 3-4	0.133	0.867
Insecure at Ages 3-4	0.122	0.878
Sociability at Ages 3-4	0.008	0.992
Behavior Problem Index Antisocial at Ages 3-4	0.405	0.595
Behavior Problem Index Anxiety at Ages 3-4	0.427	0.573
Behavior Problem Index Headstrong at Ages 3-4	0.518	0.482
Behavior Problem Index Hyperactive at Ages 3-4	0.358	0.642
Behavior Problem Index Conflict at Ages 3-4	0.336	0.664
Behavior Problem Index Antisocial at Ages 5-6	0.435	0.565
Behavior Problem Index Anxiety at Ages 5-6	0.409	0.591
Behavior Problem Index Headstrong at Ages 5-6	0.611	0.389
Behavior Problem Index Hyperactive at Ages 5-6	0.481	0.519
Behavior Problem Index Conflict at Ages 5-6	0.290	0.710
Behavior Problem Index Antisocial Ages 7-8	0.446	0.554
Behavior Problem Index Anxiety Ages 7-8	0.475	0.525
Behavior Problem Index Headstrong Ages 7-8	0.605	0.395
Behavior Problem Index Hyperactive Ages 7-8	0.497	0.503
Behavior Problem Index Conflict Ages 7-8	0.327	0.673
Behavior Problem Index Antisocial Ages 9-10	0.503	0.497
Behavior Problem Index Anxiety Ages 9-10	0.472	0.528
Behavior Problem Index Headstrong Ages 9-10	0.577	0.423
Behavior Problem Index Hyperactive Ages 9-10	0.463	0.537
Behavior Problem Index Conflict Ages 9-10	0.369	0.631
Behavior Problem Index Antisocial Ages 11-12	0.514	0.486
Behavior Problem Index Anxiety Ages 11-12	0.500	0.500
Behavior Problem Index Headstrong Ages 11-12	0.603	0.397
Behavior Problem Index Hyperactive Ages 11-12	0.505	0.495
Behavior Problem Index Conflict Ages 11-12	0.370	0.630
Behavior Problem Index Antisocial Ages 13-14	0.494	0.506
Behavior Problem Index Anxiety Ages 13-14	0.546	0.454
Behavior Problem Index Headstrong Ages 13-14	0.595	0.405
Behavior Problem Index Hyperactive Ages 13-14	0.525	0.475
Behavior Problem Index Conflict Ages 13-14	0.414	0.586

Table 2B

Percentage of Total Variance in Measurements due to Signal and Noise

Measurements of Parental Investments	%Signal	%Noise	**Measurements of Parental Investments**	%Signal	%Noise
How Often Child Goes on Outings during Year of Birth	0.329	0.671	Child Has Musical Instruments Ages 7-8	0.022	0.978
Number of Books Child Has during Year of Birth	0.209	0.791	Family Subscribes to Daily Newspapers Ages 7-8	0.023	0.977
How Often Mom Reads to Child during Year of Birth	0.484	0.516	Child Has Special Lessons Ages 7-8	0.018	0.982
Number of Soft Toys Child Has during Year of Birth	0.126	0.874	How Often Child Goes to Musical Shows Ages 7-8	0.266	0.734
Number of Push/Pull Toys Child Has during Year of Birth	0.019	0.981	How Often Child Attends Family Gatherings Ages 7-8	0.125	0.875
How Often Child Eats with Mom/Dad during Year of Birth	0.511	0.489	How Often Child is Praised Ages 7-8	0.046	0.954
How Often Mom Calls from Work during Year of Birth	0.119	0.881	How Often Child Gets Positive Encouragement Ages 7-8	0.053	0.947
How Often Child Goes on Outings at Ages 1-2	0.148	0.852	Number of Books Child Has Ages 9-10	0.013	0.987
Number of Books Child Has Ages 1-2	0.055	0.945	Mom Reads to Child Ages 9-10	0.137	0.863
How Often Mom Reads to Child Ages 1-2	0.186	0.814	Eats with Mom/Dad Ages 9-10	0.162	0.838
Number of Soft Toys Child Has Ages 1-2	0.240	0.760	How Often Child Goes to Museum Ages 9-10	0.219	0.781
Number of Push/Pull Toys Child Has Ages 1-2	0.046	0.954	Child Has Musical Instruments Ages 9-10	0.019	0.981
How Often Child Eats with Mom/Dad Ages 1-2	0.194	0.806	Family Subscribes to Daily Newspapers Ages 9-10	0.019	0.981
Mom Calls from Work Ages 1-2	0.070	0.930	Child Has Special Lessons Ages 9-10	0.015	0.985
How Often Child Goes on Outings Ages 3-4	0.123	0.877	How Often Child Goes to Musical Shows Ages 9-10	0.242	0.758
Number of Books Child Has Ages 3-4	0.012	0.988	How Often Child Attends Family Gatherings Ages 9-10	0.115	0.885
How Often Mom Reads to Child Ages 3-4	0.088	0.912	How Often Child is Praised Ages 9-10	0.036	0.964
How Often Child Eats with Mom/Dad Ages 3-4	0.170	0.830	How Often Child Gets Positive Encouragement Ages 9-10	0.041	0.959
Number of Magazines at Home Ages 3-4	0.193	0.807	Number of Books Child Has Ages 11-12	0.016	0.984
Child Has a CD player Ages 3-4	0.021	0.979	Eats with Mom/Dad Ages 11-12	0.153	0.847
How Often Child Goes on Outings Ages 5-6	0.100	0.900	How Often Child Goes to Museum Ages 11-12	0.217	0.783
Number of Books Child Has Ages 5-6	0.009	0.991	Child Has Musical Instruments Ages 11-12	0.016	0.984
How Often Mom Reads to Child Ages 5-6	0.086	0.914	Family Subscribes to Daily Newspapers Ages 11-12	0.018	0.982
How Often Child Eats with Mom/Dad Ages 5-6	0.173	0.827	Child Has Special Lessons Ages 11-12	0.013	0.987
Number of Magazines at Home Ages 5-6	0.164	0.836	How Often Child Goes to Musical Shows Ages 11-12	0.225	0.775
Child Has CD player Ages 5-6	0.015	0.985	How Often Child Attends Family Gatherings Ages 11-12	0.103	0.897
How Often Child Goes to Museum Ages 5-6	0.296	0.704	How Often Child is Praised Ages 11-12	0.026	0.974
Child Has Musical Instruments Ages 5-6	0.026	0.974	How Often Child Gets Positive Encouragement Ages 11-12	0.037	0.963
Family Subscribes to Daily Newspapers Ages 5-6	0.025	0.975	Number of Books Child Has Ages 13-14	0.023	0.977
Child Has Special Lessons Ages 5-6	0.020	0.980	Eats with Mom/Dad Ages 13-14	0.152	0.848
How Often Child Goes to Musical Shows Ages 5-6	0.304	0.696	**How Often Child Goes to Museum Ages 13-14**	0.201	0.799
How Often Child Attends Family Gatherings Ages 5-6	0.141	0.859	Child Has Musical Instruments Ages 13-14	0.015	0.985
How Often Child is Praised Ages 5-6	0.056	0.944	Family Subscribes to Daily Newspapers Ages 13-14	0.017	0.983
How Often Child Gets Positive Encouragement Ages 5-6	0.081	0.919	Child Has Special Lessons Ages 13-14	0.012	0.988
Number of Books Child Has Ages 7-8	0.007	0.993	How Often Child Goes to Musical Shows Ages 13-14	0.224	0.776
How Often Mom Reads to Child Ages 7-8	0.113	0.887	How Often Child Attends Family Gatherings Ages 13-14	0.099	0.901
How Often Child Eats with Mom/Dad Ages 7-8	0.166	0.834	How Often Child is Praised Ages 13-14	0.031	0.969
How Often Child Goes to Museum Ages 7-8	0.240	0.760	**How Often Child Gets Positive Encouragement Ages 13-14**	0.032	0.968

Table 3

The Technology for Cognitive and Noncognitive Skill Formation

Not Correcting for Measurement Error

Linear Anchoring on Educational Attainment (Years of Schooling)

No Unobserved Heterogeneity (π), Factors Normally Distributed

Panel A: Technology of Cognitive Skill Formation (Next Period Cognitive Skills)

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Self-Productivity)	$\gamma_{1,C,1}$	0.403	$\gamma_{2,C,1}$	0.657
		(0.058)		(0.013)
Current Period Noncognitive Skills (Cross-Productivity)	$\gamma_{1,C,2}$	0.218	$\gamma_{2,C,2}$	0.009
		(0.105)		(0.005)
Current Period Investments	$\gamma_{1,C,3}$	0.067	$\gamma_{2,C,3}$	0.167
		(0.090)		(0.018)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.268	$\gamma_{2,C,4}$	0.047
		(0.078)		(0.009)
Parental Noncognitive Skills	$\gamma_{1,C,5}$	0.044	$\gamma_{2,C,5}$	0.119
		(0.050)		(0.150)
Complementarity Parameter	$\phi_{1,C}$	0.375	$\phi_{2,C}$	-0.827
		(0.294)		(0.093)
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	1.601	$1/(1-\phi_{2,C})$	0.547
Variance of Shocks $\eta_{C,t}$	$\delta^2_{1,C}$	0.941	$\delta^2_{2,C}$	0.358
		(0.048)		(0.006)

Panel B: Technology of Noncognitive Skill Formation (Next Period Noncognitive Skills)

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Cross-Productivity)	$\gamma_{1,N,1}$	0.193	$\gamma_{2,N,1}$	0.058
		(0.095)		(0.014)
Current Period Noncognitive Skills (Self-Productivity)	$\gamma_{1,N,2}$	0.594	$\gamma_{2,N,2}$	0.638
		(0.090)		(0.020)
Current Period Investments	$\gamma_{1,N,3}$	0.099	$\gamma_{2,N,3}$	0.239
		(0.296)		(0.031)
Parental Cognitive Skills	$\gamma_{1,N,4}$	0.114	$\gamma_{2,N,4}$	0.065
		(0.055)		(0.015)
Parental Noncognitive Skills	$\gamma_{1,N,5}$	0.000	$\gamma_{2,N,5}$	0.000
		(0.821)		(0.203)
Complementarity Parameter	$\phi_{1,N}$	-0.723	$\phi_{2,N}$	-0.716
		(0.441)		(0.127)
Implied Elasticity of Substitution	$1/(1-\phi_{1,N})$	0.580	$1/(1-\phi_{2,N})$	0.583
Variance of Shocks $\eta_{N,t}$	$\delta^2_{1,N}$	0.767	$\delta^2_{2,N}$	0.597
		(0.076)		(0.017)

Note: Standard errors in parenthesis

Table 4

The Technology for Cognitive and Noncognitive Skill Formation

Linear Anchoring on Educational Attainment (Years of Schooling)

Allowing for Unobserved Heterogeneity (π), Factors Normally Distributed

Panel A: Technology of Cognitive Skill Formation (Next Period Cognitive Skills)

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Self-Productivity)	$\gamma_{1,C,1}$	0.479	$\gamma_{2,C,1}$	0.831
		(0.026)		(0.011)
Current Period Noncognitive Skills (Cross-Productivity)	$\gamma_{1,C,2}$	0.070	$\gamma_{2,C,2}$	0.001
		(0.024)		(0.005)
Current Period Investments	$\gamma_{1,C,3}$	0.161	$\gamma_{2,C,3}$	0.044
		(0.015)		(0.006)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.031	$\gamma_{2,C,4}$	0.073
		(0.013)		(0.008)
Parental Noncognitive Skills	$\gamma_{1,C,5}$	0.258	$\gamma_{2,C,5}$	0.051
		(0.029)		(0.014)
Complementarity Parameter	$\phi_{1,C}$	0.313	$\phi_{2,C}$	-1.243
		(0.134)		(0.125)
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	1.457	$1/(1-\phi_{2,C})$	0.446
Variance of Shocks $\eta_{C,t}$	$\delta^2_{1,C}$	0.176	$\delta^2_{2,C}$	0.087
		(0.007)		(0.003)

Panel B: Technology of Noncognitive Skill Formation (Next Period Noncognitive Skills)

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Cross-Productivity)	$\gamma_{1,N,1}$	0.000	$\gamma_{2,N,1}$	0.000
		(0.026)		(0.010)
Current Period Noncognitive Skills (Self-Productivity)	$\gamma_{1,N,2}$	0.585	$\gamma_{2,N,2}$	0.816
		(0.032)		(0.013)
Current Period Investments	$\gamma_{1,N,3}$	0.065	$\gamma_{2,N,3}$	0.051
		(0.021)		(0.006)
Parental Cognitive Skills	$\gamma_{1,N,4}$	0.017	$\gamma_{2,N,4}$	0.000
		(0.013)		(0.008)
Parental Noncognitive Skills	$\gamma_{1,N,5}$	0.333	$\gamma_{2,N,5}$	0.133
		(0.034)		(0.017)
Complementarity Parameter	$\phi_{1,N}$	-0.610	$\phi_{2,N}$	-0.551
		(0.215)		(0.169)
Implied Elasticity of Substitution	$1/(1-\phi_{1,N})$	0.621	$1/(1-\phi_{2,N})$	0.645
Variance of Shocks $\eta_{N,t}$	$\delta^2_{1,N}$	0.222	$\delta^2_{2,N}$	0.101
		(0.013)		(0.004)

Note: Standard errors in parenthesis

Table 5

The Technology for Cognitive and Noncognitive Skill Formation

Estimated Along with Investment Equation with

Linear Anchoring on Educational Attainment (Years of Schooling), Factors Normally Distributed

Panel A: Technology of Cognitive Skill Formation (Next Period Cognitive Skills)

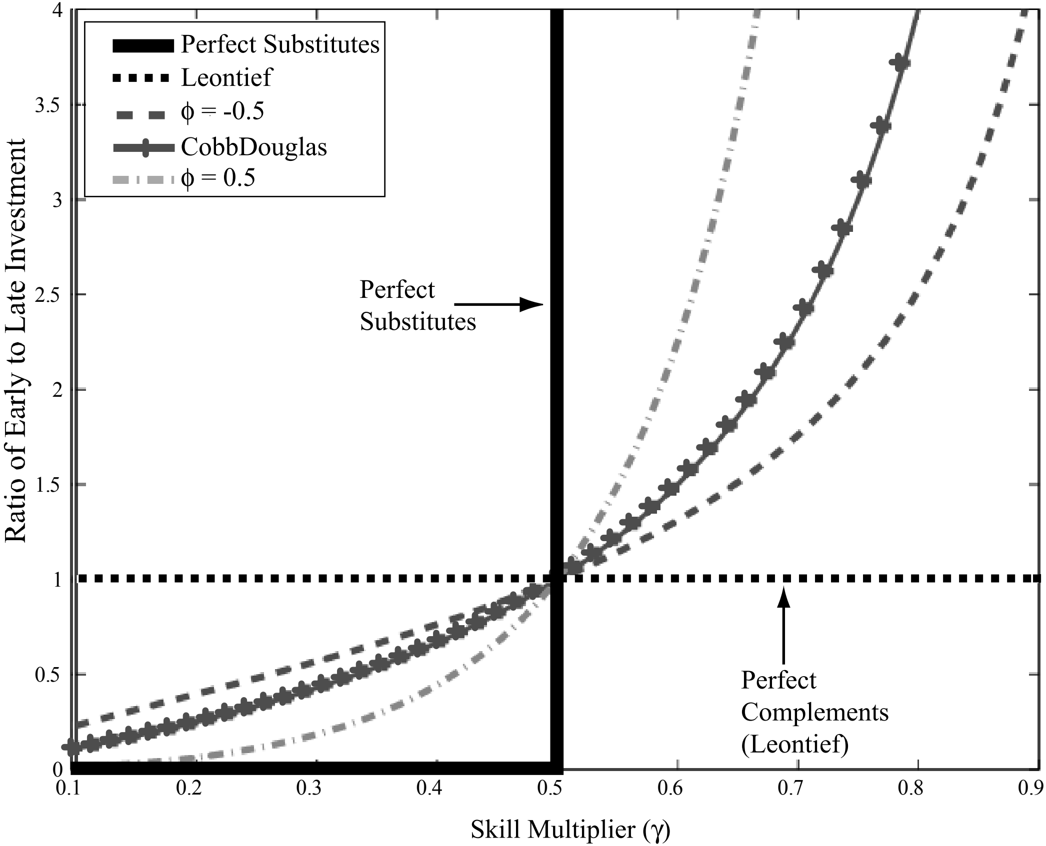
		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Self-Productivity)	$\gamma_{1,C,1}$	0.485	$\gamma_{2,C,1}$	0.884
		(0.031)		(0.013)
Current Period Noncognitive Skills (Cross-Productivity)	$\gamma_{1,C,2}$	0.062	$\gamma_{2,C,2}$	0.011
		(0.026)		(0.005)
Current Period Investments	$\gamma_{1,C,3}$	0.261	$\gamma_{2,C,3}$	0.044
		(0.026)		(0.011)
Parental Cognitive Skills	$\gamma_{1,C,4}$	0.035	$\gamma_{2,C,4}$	0.051
		(0.015)		(0.008)
Parental Noncognitive Skills	$\gamma_{1,C,5}$	0.157	$\gamma_{2,C,5}$	0.011
		(0.033)		(0.012)
Complementarity Parameter	$\phi_{1,C}$	0.585	$\phi_{2,C}$	-1.220
		(0.225)		(0.149)
Implied Elasticity of Substitution	$1/(1-\phi_{1,C})$	2.410	$1/(1-\phi_{2,C})$	0.450
Variance of Shocks $\eta_{C,t}$	$\delta^2_{1,C}$	0.165	$\delta^2_{2,C}$	0.098
		(0.007)		(0.003)

Panel B: Technology of Noncognitive Skill Formation (Next Period Noncognitive Skills)

		First Stage Parameters		Second Stage Parameters
Current Period Cognitive Skills (Cross-Productivity)	$\gamma_{1,N,1}$	0.000	$\gamma_{2,N,1}$	0.002
		(0.028)		(0.011)
Current Period Noncognitive Skills (Self-Productivity)	$\gamma_{1,N,2}$	0.602	$\gamma_{2,N,2}$	0.857
		(0.034)		(0.011)
Current Period Investments	$\gamma_{1,N,3}$	0.209	$\gamma_{2,N,3}$	0.104
		(0.031)		(0.022)
Parental Cognitive Skills	$\gamma_{1,N,4}$	0.014	$\gamma_{2,N,4}$	0.000
		(0.013)		(0.008)
Parental Noncognitive Skills	$\gamma_{1,N,5}$	0.175	$\gamma_{2,N,5}$	0.037
		(0.033)		(0.021)
Complementarity Parameter	$\phi_{1,N}$	-0.464	$\phi_{2,N}$	-0.522
		(0.263)		(0.214)
Implied Elasticity of Substitution	$1/(1-\phi_{1,N})$	0.683	$1/(1-\phi_{2,N})$	0.657
Variance of Shocks $\eta_{N,t}$	$\delta^2_{1,N}$	0.203	$\delta^2_{2,N}$	0.102
		(0.012)		(0.003)

Note: Standard errors in parenthesis

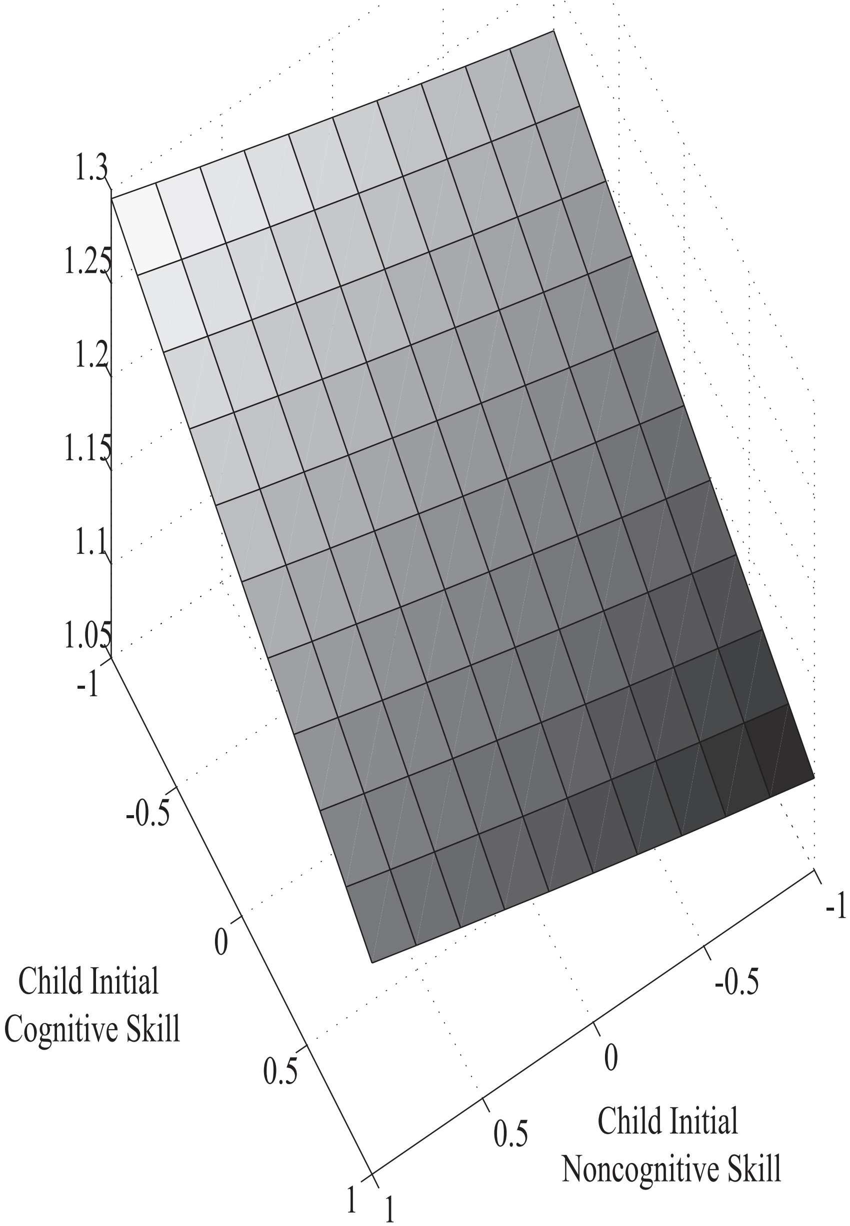
Figure 1: Ratio of early to late investment in human capital
as a function of the ratio of first period to second period investment productivity
for different values of the complementarity parameter



Note: Assumes $r = 0$.
Source: Cunha and Heckman (2007).

Figure 2

Optimal Early (Left) and Late (Right) Investments by
Child Initial Conditions of Cognitive and Noncognitive Skills
Maximizing Aggregate Education



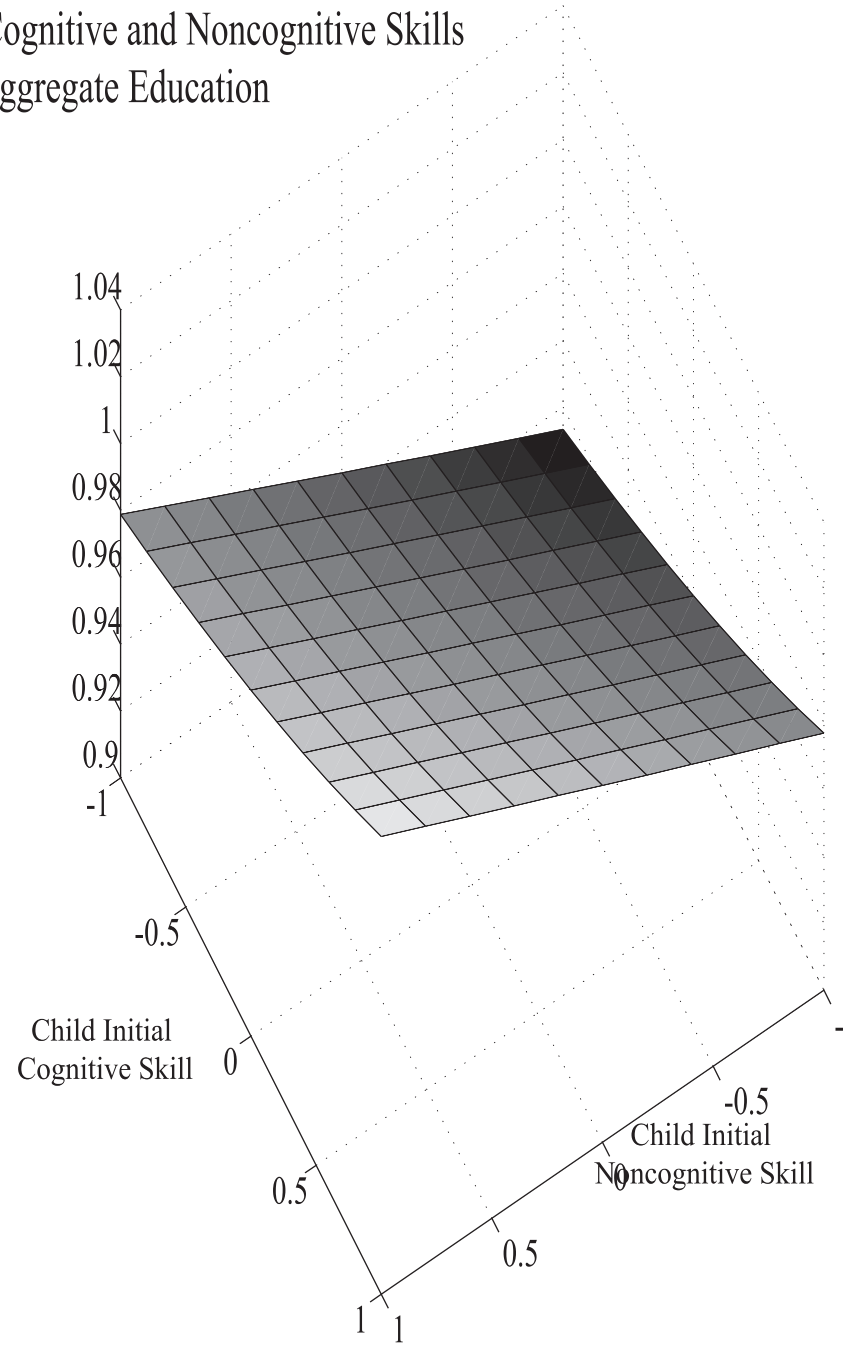


Figure 3
Optimal Early (Left) and Late (Right) Investments by
Maternal Cognitive and Noncognitive Skills
Maximizing Aggregate Education

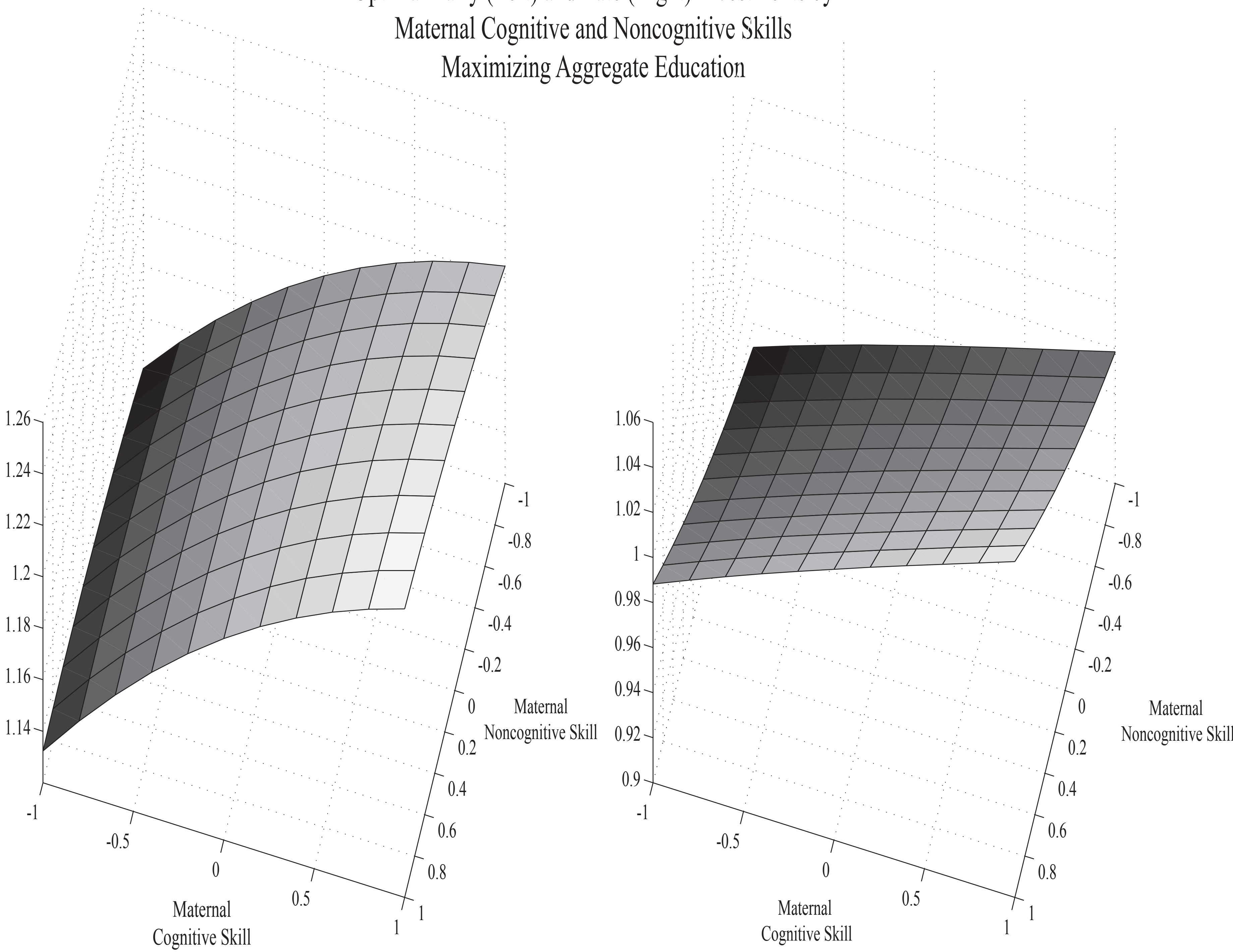


Figure 4
Ratio of Early to Late Investments by
Child Initial Condtions of Cognitive and Noncognitive Skills
Maximizing Aggregate Education (Left) and Minimizing Aggregate Crime (Right)

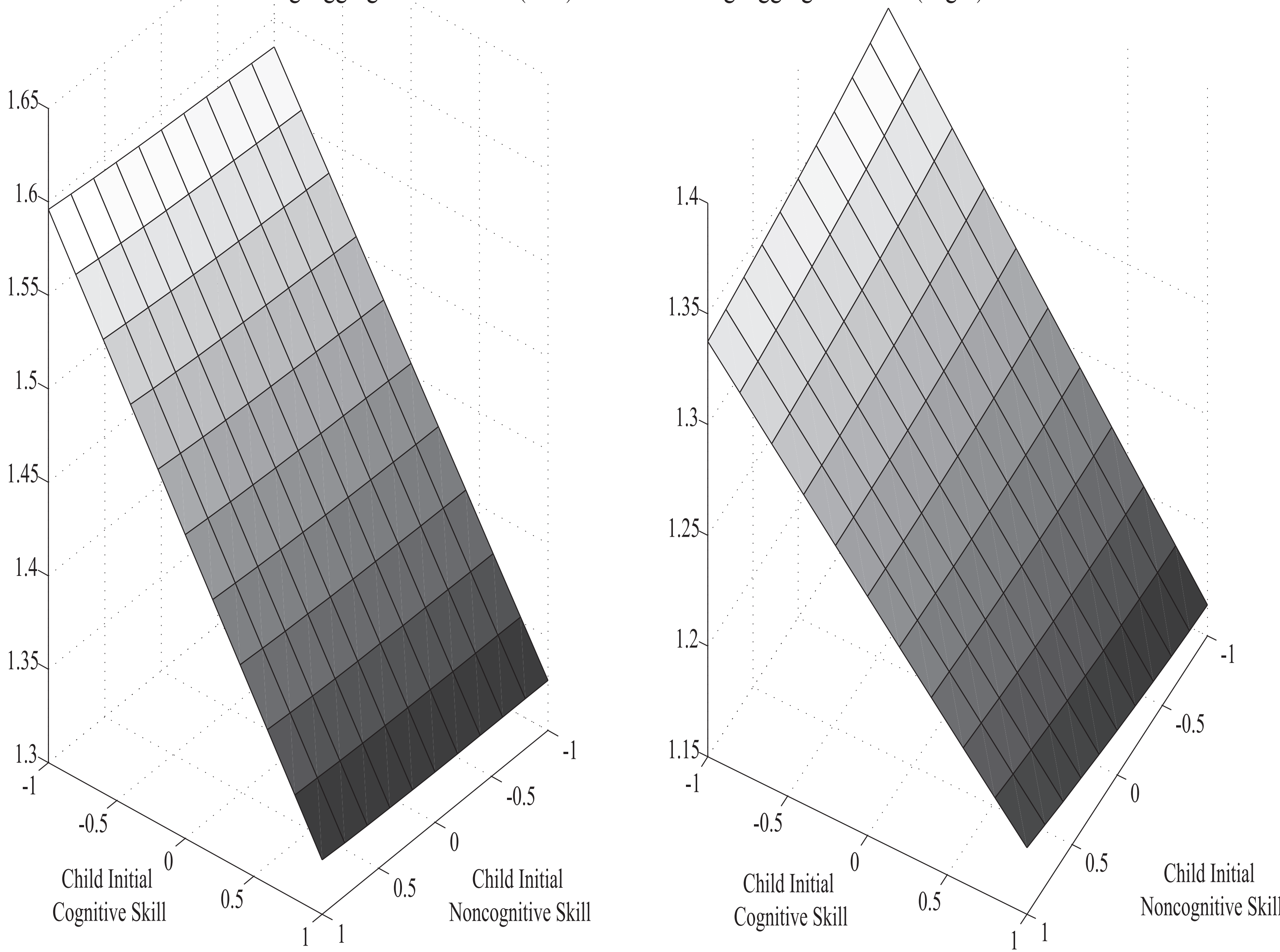


Figure 5

Ratio of Early to Late Investments by Maternal Cognitive and Noncognitive Skills

Maximizing Aggregate Education (Left) and Minimizing Aggregate Crime (Right)

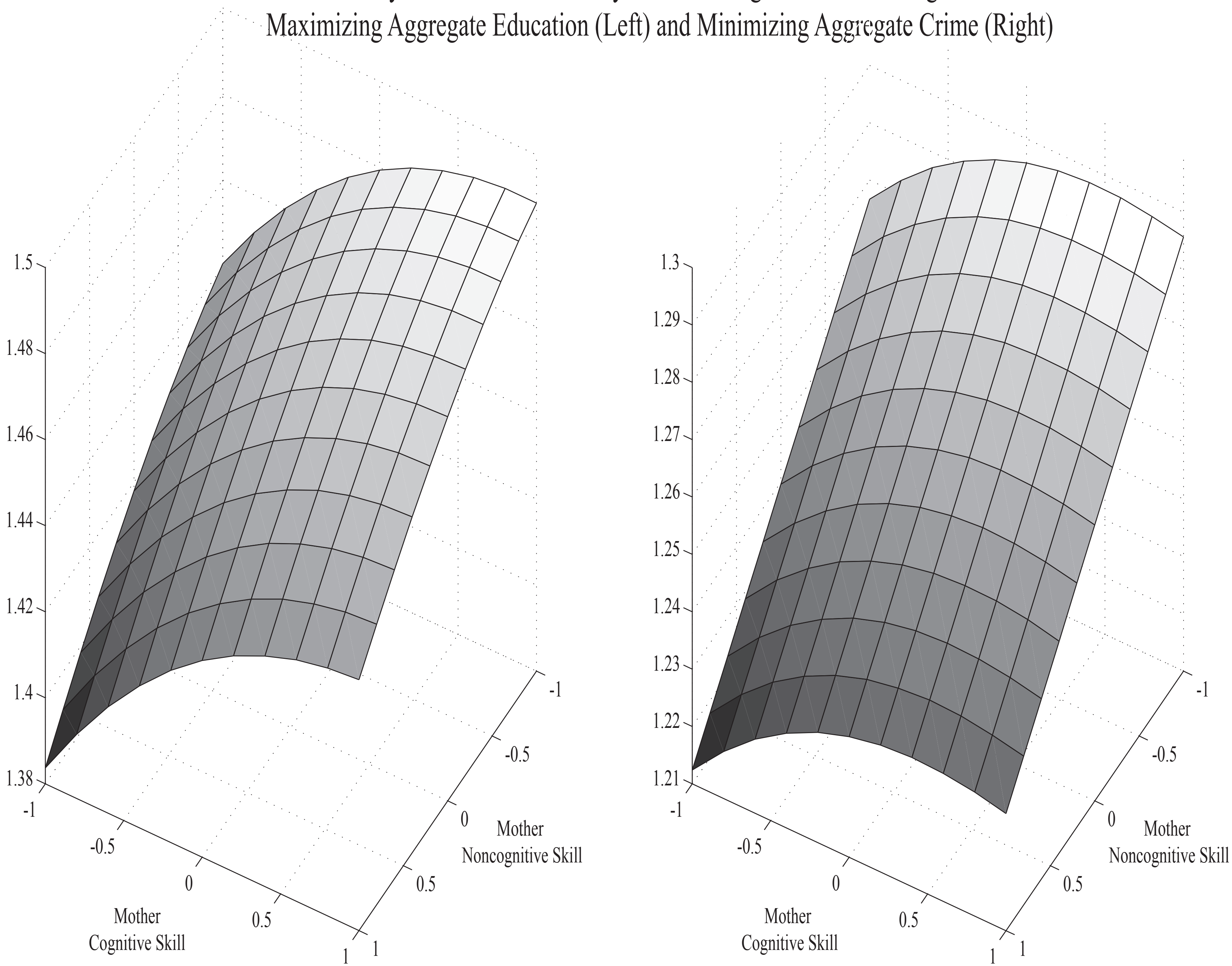


Figure 5
Densities of Ratio of Early to Late Investments
Maximizing Aggregate Education Versus Minimizing Aggregate Crime